BANOVA: An R Package for Hierarchical Bayesian ANOVA

Chen Dong
University of Maryland, College Park

Michel Wedel
University of Maryland, College Park

Abstract

In this paper, we develop generalized hierarchical Bayesian ANOVA, to assist experimental researchers in the behavioral and social sciences in the analysis of the effects of experimentally manipulated within- and between-subjects factors. The method alleviates several limitations of classical ANOVA, still commonly employed in those fields. An accompanying R Package for hierarchical Bayesian ANOVA is developed. It offers statistical routines and several easy-to-use functions for estimation of hierarchical Bayesian ANOVA models that are tailored to the analysis of experimental research. Markov chain Monte Carlo (MCMC) simulation is used to simulate posterior samples of the parameters of each model specified by the user. The core program of all models is written in R and JAGS. After preparing the data in the required format, users simply select an appropriate model, and estimate it without any advanced coding. The main aim of the R package is to offer freely accessible resources for hierarchical Bayesian ANOVA analysis, which makes it easy to use for behavioral researchers.

Keywords: hierarchical Bayes regression, ANOVA, Markov chain Monte Carlo, BUGS modeling.

1. Introduction

Because of its computational attractiveness and the ease of interpretation of its statistical tests and the corresponding tables of means, ANOVA (Fisher 1921, 1925) is implemented in most statistical packages and continues to garner tremendous popularity in applied research. As it is about to see its centennial, especially the behavioral sciences still rely heavily on ANOVA for the analysis of their data from experiments with human subjects (e.g. Cardinal and Aitken 2005). Yet, the standard approach to ANOVA is based on several assumptions that often do not hold for the data typically collected in those fields of research.
First, ANOVA assumes a continuous homoscedastic i.i.d. Normal distributed dependent variable. Categorical variables, however, for which ANOVA has long known not to be appropriate (Cochran 1940), abound in the behavioral sciences. These include measurements of perceptions, attitudes and intentions on categorical rating and multiple choice scales, and binary and count measures of human attention, memory and decision making (Nunnaly 1967; Thorndike 1971; Lord and Novick 1968). In addition, continuous measures such as response times, which often have high skewness or kurtosis, are commonly used as measures of human behavior. The distributional properties of most of these variables violate the assumptions underlying ANOVA, and extensive research into the effect of these violations has shown that they may lead to both excess type-I and type-II errors in significance testing (Ito 1980; Tan 1982; Tiku 1971). Transformations of the data, such as the log and square-root (for counts), logit and arcsine (for proportions), rank (for ordered categorical variables), and Box-Cox transformations (for various measures) have been used as a way to render the empirical distribution closer to the Normal (Bartlett 1947; Box and Cox 1964; Draper and Hunter 1969; Conover and Iman 1976). These transformations, however, often do not provide a satisfactory solution, because they may cause the ANOVA tables of means to lose some of their appealing interpretations, while significance levels of the transformed and original data do not necessarily correspond. But, modern statistical solutions are available in the form of Generalized Linear Models (McCullagh and Nelder 1989). The application of GLMs capitalizes on the fact that ANOVA is a special case of linear regression models, and GLMs extend those to a wide variety of distributions of the dependent variable in the exponential family. However, for applied researchers a downside of the use of GLM to analyze designed experiments that involve of multiple factors and interactions that need to be represented in the model through dummy variables, is that the interpretation of estimates of coefficients of these dummy variables is not as easy as interpreting the output of ANOVA, and indeed, may often be quite cumbersome.

Second, in the behavioral sciences experiments often employ a combination of between- and within-subjects factors, leading to nested and repeated measurement designs for which established (split-plot and repeated measures) ANOVA procedures are available in most statistical packages. These mixed ANOVA models may have both fixed and random effects (Hartley and Rao 1967; Scheffé 1957), but assume a balanced design, a continuous dependent variable and categorical independent variables. Unbalanced designs, unequally spaced measurements, and continuous covariates such as encountered in ANCOVA, violate these assumptions. As a consequence, in certain fields of behavioral science some experimental behavioral science researchers have resorted to hierarchical linear models (Breslow and Clayton 1993; Longford 1987; Raudenbusch 1988; Raudenbush 1999), in particular for the analysis of quasi experiments. These models, which take on a variety of forms and go under different names in the literature, are special cases of hierarchical Bayes models (Lindley and Smith 1972; Press 2003; Gelman, Carlin, Stern, and Dunson 2013). They allow for more general covariance structures and data hierarchies than repeated measures ANOVA. Gelman (2005) argued the importance of hierarchical Bayes formulations of ANOVA, and showed how the principles of ANOVA are helpful in understanding hierarchical linear models. In addition, hierarchical Bayes models can accommodate non-Normal dependent variables that render the application of classical ANOVA and hierarchical linear models problematic. The Bayesian approach in addition offers a number of theoretical and pragmatic advantages as a framework for inference and testing that have been widely acknowledged (Bernardo J 2000; Press 2003; Savage 1954). Indeed, in behavioral research several advantages of Bayesian inference are increasingly recognized,
in that it provides inferences based on finite samples, avoids pitfalls of classical hypothesis
testing, and may not only reject but also support hypotheses (Kruschke 2013; Rouder, Speck-
man, Sun, Morey, and Iverson 2009). Hierarchical Bayes models can now be relatively easily
be implemented using existing statistical software, such as BUGS and JAGS (Lunn, Thomas,
Best, and Spiegelhalter 2000).

Nevertheless, applied experimental researchers in the behavioral sciences continue to resort
to standard ANOVA in many cases in spite of these limitations and in spite of the availability
of these modern superior alternatives, because of its ease of application and interpretation,
widespread availability in standard statistical packages, lack of familiarity with better alter-
natives, and/or the effort involved in programming alternative methods or interpreting their
output. The present paper attempts to help remedy this undesirable state of affairs by develop-
ing an R package for hierarchical Bayes ANOVA that addresses the most salient limitations
of classical ANOVA, yet is easy to use and retains many of the familiar features of the outputs
of classical ANOVA. It deals with a wide range of distributions for the dependent variable,
with hierarchical data structures and between- and within- subjects design factors, as well as
continuous covariates.

The key insight behind the approach is that ANOVA and ANOCOVA are special cases of
linear regression and that once an ANOVA model is formulated as a hierarchical linear model,
subject-level parameters become incidental and inference focuses entirely on the population-
level model, which is where main effects and interactions of within- and between-subjects
factors are represented and tested. Assuming that lower level parameters describing subject
heterogeneity follow Normal distributions, it follows that significance tests of main effects and
interactions, variance decompositions and tables of means can be computed in a similar way
as they would be in standard ANOVA. This then allows for the analysis of dependent variables
with a wide variety of distributional forms with hierarchical models, but at the same time
retaining much of the appealing output from standard ANOVA for experimental data. The
underlying estimation methods are Markov Chain Monte Carlo algorithms implemented in the
JAGS software. The user of the package needs to input the data and set up a few parameters.
The package then sets up a JAGS program and analyses the data with a hierarchical Bayes
ANOVA, using MCMC estimation. JAGS was chosen as an interface, because after calling the
R package in addition to the estimation results, the JAGS code will be available for inspection
and modification by the behavioral researcher. Importantly, although the underlying models
are Hierarchical Bayes models, the output of these models is presented in a form that is very
familiar to users of standard ANOVA, including (Bayesian) p-values and effect sizes, and
tables of means with confidence intervals.

The remainder of this paper is organized as follows. In section 2, we discuss the hierarchical
Bayesian approach to ANOVA. In section 3, the architecture and tutorial of the R package is
discussed. Bayesian estimation of parameters and other quantities of interest are introduced
as well. We include examples in section 4. The last section gives the conclusion and future
work.

2. Bayesian ANOVA Models

We assume data are collected in an experiment in which a sample of subjects have partic-
ipated, and have been subjected to between-subject as well as within-subject experimental manipulations. Repeated measurements of one or more dependent variables are taken on each subject, while continuous or categorical covariates may have been measured. The hierarchical Bayesian approach to ANOVA then consists of two sub-models: level-1—the subject level, and level-2—the population level. The subject-level model represents the effects of within-subject factors and covariates, and the population-level model represents the influence of between-subject level factors. The population-level model expresses the ANOVA of interest.

In the subject-level model, each outcome of the dependent variable, \( y_i \), with \( i \) indexing data points, is assumed to be generated from a particular distribution in the exponential family, \( f(y_i|\mu_i) \) (and even other distributions can be accommodated). The mean, \( \mu_i \), of the distribution depends on the independent variables through a suitable link function \( g(\cdot) \) (McCullagh and Nelder 1989). The within-subject factors and their interactions are indexed by \( p(p = 1, 2, ..., P) \). Each index \( p \) represents a batch of \( J_p \) coefficients: \( \beta^p \).

\[
\eta_i = \sum_{p=0}^{P} \sum_{j=1}^{J_p} X^p_{ij} \beta^p_{j,s_i},
\]

where \( s_i \) is the subject index of data point \( i \).

The population-level model allows for unobserved heterogeneity among subjects, because the subject-level coefficients \( \beta^p \) are assumed to follow a multivariate normal distribution. The between-subject factors and their interactions are indexed by \( q(q = 1, 2, ..., Q) \); \( q = 0 \) denotes the constant term. Then, using the notation in Gelman (2005), the population-level ANOVA can be written as:

\[
\beta^p_{j,s} = \sum_{q=0}^{Q} \theta^{p,q}_{j,k,s} + \delta^p_{j,s},
\]

Each \( q \) represents a batch of \( K_q \) coefficients: \( \theta^{p,q}_{j,k,s} \). \( k = 1, ..., K^q \); \( K^q_s \) indexes coefficient \( k \) in batch \( q \) corresponding to the treatment of subject \( s \). For example, in the simple case of one 3-level within-subject factor \( D(P = 2, J_1 = 1 \text{ and } J_2 = 3) \) and two 2-level between-subject factors \( A \) and \( B \), and the \( AB \)-interaction, \( Q = 3 \), equation (2) and equation (3) reduce to (with the parameter of the last level of each factor set to zero):

\[
\eta_i = \beta^0_{1,s} + \sum_{j=1}^{2} X^D_{ij} \beta^D_{j,s_i},
\]

\[
\beta^0_{1,s} = \theta^0 + \theta^A_{1,k^A} + \theta^B_{1,k^B} + \theta^{AB}_{1,k^A,k^B},
\]

\[
\beta^D_{1,s} = \theta^D_{1} + \theta^{DA}_{1,k^A} + \theta^{DB}_{1,k^B} + \theta^{DAB}_{1,k^A,k^B},
\]

\[
\beta^D_{2,s} = \theta^D_{2} + \theta^{DA}_{2,k^A} + \theta^{DB}_{2,k^B} + \theta^{DAB}_{2,k^A,k^B}.
\]
Here, equation (4b) contains the overall intercept ($\theta_0$) and the main effects of the between-subject factors $A$ ($\theta_{A,k}$) and $B$ ($\theta_{B,k}$) and their interaction ($\theta_{AB}$). Equation (4c) contains the main effect of the (first level) of the within-subject factor $D$ ($\theta_D$), and its’ two- and three-way interactions with the between-subject factors and $B$. Similarly, equation (4d) contains the main effect of the second level of $D$, and its interactions with $A$ and $B$.

The population-level ANCOVA model can be expressed as a linear model with a design matrix $Z$ that contains all between-subject factors and their interactions (and a constant term):

$$
\beta_{j,s} = \sum_{k=1}^{K} Z_{s,k} \theta_{j,k} + \delta_{j,s},
$$

where $Z_{s,k}$ is an element of $Z$, a $S \times K$ matrix of covariates, and $K$ is the number of parameters. $\theta_{j,k}$ is a hyper-parameter which captures the effects of between-subject factor $q$ on the parameter $\beta_{j,s}$ of within-subject factor $p$. The error $\delta_{j,s}$ is assumed to be normal: $\delta_{j,s} \sim N(0, \sigma^{-2})$. Proper, but diffuse priors are assumed: $\theta_{j,k} \sim N(0, s)$, and $\sigma_p \sim Gamma(a, b)$, where $s, a, b$ are hyper-parameters.

The Hierarchical Bayes ANOVA model is estimated capitalizing on the fact that it is a special case of hierarchical generalized linear models, that is, using equations (1), (2) and (5). We use effects-coding of the factors in the matrices $X$ and $Z$ (Overall, Spiegel, and Cohen 1975). It is important to note that equation (3) is the equation that is of key interest for inference. It contains the parameters that specify the population-level ANOVA model. It is as if the subject-level coefficients $\beta_{j,s}$ are the (Normally distributed) "dependent variables" in an ANOVA, specified by the between-subject factors in equation (3). Thus, inferences focuses on the parameters in equation (3). We first specify the specific outcome variables that are accommodated in the R package below.

**Continuous responses:** To model continuous data, a normal distribution can be assumed for $y_i$:

$$
y_i = \eta_i + \epsilon_i; \quad \epsilon_i \sim N(0, \sigma^{-2}),
$$

where $\eta_i$ is defined in equation (2), and the scale parameter $\sigma$, $\sigma \sim Gamma(\alpha, \beta)$.

To describe data with "outliers" or fatter tails than the normal, the distribution of $\epsilon_i$ is assumed to follow a $t$-distribution, with an unknown number of degrees of freedom, assumed to follow a Poisson distribution:

$$
\epsilon_i \sim t(0, \phi^{-2}, \nu), \phi \sim Gamma(\alpha, \beta), \nu \sim Poisson(\lambda),
$$

where $\alpha, \beta, \lambda$ are hyper-parameters.

**Binary responses:** To model data $y_i$ that take on the values 0 and 1, a Bernoulli distribution is assumed,

$$
y_i \sim Binomial(1, p_i), \quad p_i = \text{logit}^{-1}(\eta_i),
$$

where $\text{logit}(x) = \ln \frac{x}{1-x}$ is the standard logit link-function.

If the data $y_i$ represents the number of successes in a sequence of $B$ independent Bernoulli experiments, then,

$$
y_i \sim Binomial(B, p_i), \quad p_i = \text{logit}^{-1}(\eta_i),
$$
**Count responses**: To model count data $y_i$ that can take on values in $0, 1, 2, ...$, the Poisson distribution is assumed:

$$y_i \sim \text{Poisson}(\lambda_i); \lambda_i = \exp(\eta_i).$$  \hspace{1cm} (10)

**Ordered categorical responses**: To model data $y_i$ that are ordered categorical and can take on the values $1, ..., M$, an ordered logistic model is used,

$$Pr(y_i > m) = \text{logit}^{-1}(\eta_i - c_{m-1}); m = 1, ..., M - 1.$$  \hspace{1cm} (11)

The cut-point parameters $c_k$ are constrained: $0 = c_1 < c_2 < \cdots < c_{K-1}$. We assume $c_1 = 0$, and the other cut-points follow a Normal distribution $c_m \sim N(0, \bar{\sigma}_m^2)$, with $\bar{\sigma}_m^2 \sim \text{Uniform}(0, 100)$.

**Multinomial responses**: To model data $y_i$ that are categorical and can take on the values $1, ..., K$, a multinomial logistic model (MNL) is used,

$$Pr(y_i = k) = \frac{\exp(\eta_{i,k})}{\sum_{k=1}^{K} \exp(\eta_{i,k})}$$  \hspace{1cm} (12)

where $\eta_{i,k} = \sum_{p=0}^{P} \sum_{j=1}^{J_p} X_{i,j}^k \beta_{p,j}$, and $X_{i,j}^k$ is the design matrix corresponding to each response category $k (k = 1, ..., K)$ of $y_i$.

### 3. BANOVA R package

**3.1. Obtaining the software**

The BANOVA package is an add-on package to the statistical software R. It is free and can be downloaded from the Comprehensive R Archive Network (CRAN, http://CRAN.R-project.org/...). The base of BANOVA package is implemented in R and JAGS. Thus, an additional system requirement is the JAGS software, which can be freely downloaded from (http://mcmc-jags.sourceforge.net). The package also imports two other packages, runjags (Denwood 2013) and coda (Plummer, Best, Cowles, and Vines 2006) in order to connect R and JAGS and perform convergence diagnostics. Note, the above two imported packages do not necessarily need to be installed before installing BANOVA. They are automatically attached to the package and loaded when the package BANOVA is loaded. However, the JAGS software must be installed in order to estimate any of the models introduced above. The package will automatically detect the location of JAGS software and connect it via runjags. Once R and JAGS have been installed, BANOVA can be loaded using the following command,

```R
R> library('BANOVA')
```

**3.2. Functionality**

BANOVA can fit the Bayesian hierarchical ANOVA models introduced in the previous section. As explained there, the response data follows a wide variety of distributions including Normal,
student’s t, Poisson, binomial, ordered and unordered Multinomial distributions. Each of the corresponding models can be fitted by a specific function in the package. The names of these functions have the form of ‘BANOVA.Bin’, where the first part specifies the general name and the second part after the ‘.’ specifies the form of the likelihood. Currently, there are seven models included in the package.

1. `BANOVA.Bern()` — the model in which the response variable follows a Bernoulli distribution (equation 8).
2. `BANOVA.Bin()` — the model in which the response variable follows a binomial distribution (equation 9).
3. `BANOVA.Multinomial()` — the model in which the response variable follows an unordered Multinomial distribution (equation 12).
4. `BANOVA.Normal()` — the model in which the response variable follows a Normal distribution (equation 6).
5. `BANOVA.ordMultinomial()` — the model in which the response variable follows an ordered categorical distribution (equation 11).
6. `BANOVA.Poisson()` — the model in which the response variable is considered a count variable which follows a Poisson distribution (equation 10).
7. `BANOVA.T()` — the model in which the response variable follows a t distribution (equation 7).

The predictor for each Bayesian ANOVA model is specified as a regular R object, which is similar to the `lm()` and `glm()` objects in R. This means that the `summary()`, `print()` and `predict()` functions can be applied to the object in question after fitting the model. In addition to the common R object functions, the package also includes several useful functions such as `conv.diag()`, `table.means()`, `table.pvalues()` and so on. Their use is illustrated in following sections.

### 3.3. Data input

When the data is in `.csv` or other formats, it can be loaded with the R function `read.csv()` or other import functions. The package expects the data imported to be in a long format where each row corresponds to one trial, replication, or time point per subject. Thus, each subject will have data in multiple rows. Subject ID values must be included in the data (see Figure 1), and the other columns in the data set contain the dependent variable(s), the covariates, and the between- and within- subject experimental factors. The between-subject variables, which are constant within each subject, will have the same value in all rows containing the data for one subject. The attribute of each of the factors must be specified as one of the following three classes: "integer", "numeric" or "factor". The function `class()` in R can be used to check the classes of factors. For example,

```r
R> class(x) # will display 'integer' number of classes in variable x
R> x <- as.factor(x) # class of x is changed to 'factor'
```
Figure 1: Example of the Format of Input Data

For the multinomial response model (equation 12), the data format is somewhat more complex. The within-subject data for each subject must be stored in one item of a 'list' in R. For example, if there are 100 subjects, then the list must contain 100 items where each item includes multiple rows that denote the values of the within-subject variables. The between-subject data is stored in a separate data frame where each row corresponds to one subject. The order of the between-subject data must match the order of within-subjects data. For example, although the 'choicedata' of BANOVA package is already in a long format, both within-subject and between-subject data needs to be further manipulated. The following R code can be used for that purpose:

```R
R> data(choicedata)
R> #generate within-subject data (convert the within-subjects variables to a list)
R> dataX <- list()
R> for (i in 1:nrow(choicedata)){
R>   logP <- as.numeric(log(choicedata[i,3:8]))
R>   dataX[[i]] <- as.data.frame(logP) - mean(logP) # mean center logP
R> }
R> # generate between-subject data
R> dataZ <- choicedata[,9:13]
```
3.4. Estimation of the coefficients

As explained above, the Hierarchical Bayes ANOVA model is estimated using the fact that it is a special case of a hierarchical generalized linear model, that is, using equations (1), (2) and (5). The conditional posterior distribution, denoted by $\pi(\cdot)$, of the parameters $\beta_{j,s}^p$ is obtained from the likelihood and priors:

$$
\pi(\beta_{j,s}^p | \cdot) \propto \pi(y_i | \beta^1, ..., \beta^p, x_i) \pi(\beta^1 | \theta^1, Z) \cdots \pi(\beta^p | \theta^p, Z) \pi(\theta^1) \cdots \pi(\theta^p),
$$

(13)

where $\pi(y_i | \beta^1, ..., \beta^p, x_i)$ is the likelihood determined by equations (1), (2) and (6) to (10); $\beta^p = (\beta^p_1, ..., \beta^p_J_p)$, $p = 1, ..., P$, with $\beta^p_j = (\beta^p_{j,1}, ..., \beta^p_{j,s})$, $j = 1, ..., J_p$, are the coefficients of factor $p$; $\theta^p = (\theta^p_{j,k,1}, \theta^p_{j,k,2}, ..., \theta^p_{j,k,Q})$, $j = 1, ..., J_p$, $k = 1, ..., K_q$ are the population-level parameters; $\pi(\theta^p)$, $p = 1, ..., P$ is the prior of population-level parameters $\theta^p$; and $\pi(\beta^p | \theta^p, Z)$, $p = 1, ..., P$, is the prior determined by the population model in equation (3). We are interested in the effects of between-subject level factors, captured by $\theta^p_{j,k}$, $k = 1, ..., K_q$ using the notation in equation (3). The conditional posterior distribution of the parameters $\theta^p_{j,k}$ is:

$$
\pi(\theta^p_{j,k} | \cdot) \propto \pi(\beta^p_j | \theta^p, Z) \pi(\theta^p)
= \pi(\beta^p_j | \theta^p, Z) \prod_{j,k} \pi(\theta^p_{j,k,1}) \prod_{j,k} \pi(\theta^p_{j,k,2}) \cdots \prod_{j,k} \pi(\theta^p_{j,k,Q}).
$$

(14)

where each $\theta^p_{j,k}$ is assumed to follow a normal prior.

3.5. Starting values and burn-in period

Successful implementation of MCMC algorithm requires proper starting values and a sufficiently long burn-in period to make sure the convergence of chains. For the burn-in period, typically the first 1000 to 5000 draws are discarded. Users can easily adjust that number in the arguments of all BANOVA.*() functions. Because the hierarchical Bayesian ANOVA models usually involves many hyper parameters, the starting values of all parameters are assigned by the R package. Specifically, the random parameters without any constraint are assigned values drawn from a Normal distribution using the rnorm() function (the starting values of $\theta^p_{j,k}$ in equation (14), for example). For those parameters with constraints, for instance, the hyper parameters $\alpha, \beta, \nu$ in equation (7), are assigned fixed values ($\alpha = 1, \beta = 1, \nu = 1$). These values can also be found through the corresponding JAGS code.

3.6. JAGS code

Based on equation (13) and (14), the models are built and estimated in JAGS (Just Another Gibbs Sampler, Plummer (2003)). The BANOVA package generates the JAGS code fully automatically, and runs it. The JAGS software takes care of the work involved in estimating model parameters by constructing a MCMC algorithm to sample from the posterior distributions. The JAGS program allows users to write their own models and prior distributions and frees them from dealing with the implementation details of different models and samplers. The JAGS code is produced as part of the output of our package, so that users can inspect and modify it. The following R command provides users with the JAGS code generated for the model in question:
3.7. Convergence diagnostics

There is a large number of convergence diagnostics available (e.g. Gill 2007). In the output of the package, two convergence diagnostics are reported: the Geweke diagnostic (Geweke 1991), and the Heidelberg and Welch (Heidelberger and Welch 1983) diagnostic. These two convergence diagnostics are calculated based on only a single MCMC chain, which saves some computation time and is less cumbersome for the applied user. Both diagnostics require a single chain and may be applied with any MCMC method. The functions `geweke.diag` and `heidel.diag` in R package `coda` are incorporated in our package and used to compute the convergence diagnostics. If so desired the user can apply other diagnostics from the `coda` package manually.

Geweke’s convergence diagnostic is calculated by taking the difference between the means from the first $m_A$ iterations and the last $m_B$ iterations, where $m$ is the total number of iterations. If the ratios $m_A/m$ and $m_B/m$ are fixed and $m_A + m_B < m$, then by the central limit theorem, the distribution of this diagnostic approaches a standard normal as $m$ tends to infinity. In our package, $m_A = 0.1m$ and $m_B = 0.5m$.

The Heidelberg and Welch diagnostic is based on a test statistic to accept or reject the null hypothesis that the Markov chain is from a stationary distribution. The present package reports the Cramer-von Mises statistic to test for stationarity. The hypothesis test is based on Brownian bridge theory where the sequence of iterates is from a stationary. The test is iteratively applied on batches of draws from the posterior distributions. If the null hypothesis is rejected, the first 10% of the iterations are discarded and the stationarity test repeated. If the test fails again, an additional 10% of the iterations are discarded and the test is repeated. The process continues until 50% of the iterations have been discarded and the test still rejects. Our package uses the function `heidel.diag` in the `coda` package and sets the parameters $\epsilon = 0.1$, $pvalue = 0.5$.

To obtain the convergence tests, the following R command is used:

\begin{verbatim}
R> conv.diag(res) // res is a fitted object from any of the models
\end{verbatim}

3.8. Tables of means

One key output of the package is a table of means for the categories of the factors at both level 1 and level 2. As explained above, we use effects coding to estimate the parameters of categorical variables, using the last level of each factor as the reference level. However, especially when there are multiple factors and interactions, interpretation of the parameter estimates is cumbersome. Therefore, posterior samples of each $\theta_{j,k}$ in equation (3) generated by MCMC are used in the calculation of ‘tables of means’, similar to those produced by standard ANOVA. The advantage of doing this is that this output is familiar to behavioral researchers and relatively easy to interpret. Because these statistics are computed for each draw from the posterior distribution of the parameters, statistics from their posterior distributions are
readily available. Specifically, the package computes statistics of interest such as 95% credible intervals and posterior standard deviations.

Let $\theta^{pq}_{j,k,m}$ denote the posterior sample of $\theta^{pq}_{j,k}$ in $m$th iteration of the MCMC chain. Then the grand mean is:

$$\bar{\mu} = \frac{\sum_{m} g^{-1}(\theta^{00}_{m})}{M},$$  

where $\theta^{00}_{m}$ is the $m$th draw of the level-2 intercept corresponding to the level-1 intercept, which is equal to the grand mean. The 95% credible interval is simply provided by the 2.5% and 97.5% quantiles of the posterior distribution of $\{g^{-1}(\theta^{00})\}$, and the posterior standard deviation is also computed from the draws of the posterior distribution. Higher order tables are computed as illustrated below.

For example, in computing the one-way table of means of a level-1 factor $A$, the posterior mean of its level $j$ is calculated as:

$$\bar{\mu}^A_j = \frac{\sum_{m} g^{-1}(\theta^{00}_{m} + X^A_j \theta^A_{0m})}{M},$$  

where $X^A_j$ is the effects coded column vector of factor levels, corresponding to level $j$ of factor $A$, and in which all other factors and covariates are set to be 0; $\theta^A_{0m}$ is the estimated vector of level-2 intercepts corresponding to level-1 factor $A$.

For another example, in computing the one-way table of means of a level-2 factor $B$, the posterior mean of its level $j$ is calculated as:

$$\bar{\mu}^B_j = \frac{\sum_{m} g^{-1}(\theta^{00}_{m} + Z_{j,B}^T \theta^{0B}_{0m})}{M},$$  

where $Z_{j,B}$ is the effects coded column vector of factor levels, corresponding to level $j$ of factor $B$, and in which all other factors and covariates are set to be 0; $\theta^{0B}_{0m}$ is the estimated vector of level-2 coefficients of the effect of factor $B$ on the level-1 intercept.

Continuing the example, the means of the two-way table classified by $A$ and $B$ (level $j$ of factor $A$ and level $k$ of factor $B$) is calculated as:

$$\bar{\mu}^{AB}_{j,k} = \frac{\sum_{m} g^{-1}(\theta^{00}_{m} + X^A_j \theta^{AB}_{j,k,m} Z^B_{k,m})}{M},$$  

where $\theta^{AB}_{j,k,m}$ is the $m$th draw from the coefficient matrix, the $k$th row of which is a vector of level-2 coefficients representing the effect of factor $B$ on to the $j$th level of level-1 factor $A$.

Based on the above formulas, the function `table.means()` computes the tables of means (currently limited to two-way interactions at each level of the model) and their posterior quantiles.

R> table.means(res) // res is a fitted object from any of the models

3.9. Table of sums of squares and effect sizes

If the experimental design is balanced, a variance decomposition can be performed at the level-2 equations (3) or (5) to produce information on sums of squares and effect sizes. Both of
these sets of statistics are important in interpreting the results of experiments in behavioral research. For this purpose, it is convenient to consider the ANOVA as a regression as in equation (5), so that for \( \theta^p_j = (\theta^p_{j,1}, \theta^p_{j,2}, \ldots, \theta^p_{j,K}) \), the total sum of squares can be represented as

\[
SS(\theta^p_j) = \theta^p_j Z^p_j \beta^p_j.
\] (19)

In the package, equation (23) is estimated as

\[
SS(\theta^p_{pq,j}) = \sum \theta^p_{pq,j,m} Z^p_j \beta^p_{pq,j,m} M
\]

where \( m \) indexes the posterior samples of \( \theta^p_{pq,j} \) and \( Z^p_j \beta^p_{pq,j} \). If the design matrix \( Z \) is orthogonal, then the sum of squares attributable to each of the factors and their interactions can be written in terms of the submatrix \( Z_q \) of \( Z \) corresponding to each factor \( q \) and its coefficients \( \theta^p_{pq,j} = (\theta^p_{pq,j,1}, \ldots, \theta^p_{pq,j,K_q}) \), \( q = 1, \ldots, Q \). (e.g. Draper and Smith 1998),

\[
SS(\theta^p_{pq,j}) = \theta^p_{pq,j} Z^p_j \beta^p_{pq,j},
\] (20)

Where \( \theta^p_{pq} \) is the vector of coefficients of the dummy variables corresponding to factor or interaction \( q \). Equation (24) is estimated as

\[
SS(\theta^p_{pq,j}) = \sum \theta^p_{pq,j,m} Z^p_q \beta^p_{pq,j,m} M
\]

in the package. For orthogonal designs it holds that

\[
SS(\theta^p_j) = SS(\theta^p_{p1,j}) + SS(\theta^p_{p2,j}) + \cdots + SS(\theta^p_{pQ,j}).
\] (21)

If the design is not balanced, type III sum-of-squares are computed. These reflect the presence of a main effect after the other main effects and interactions are accounted for, and are valid in the presence of significant interactions (Fox 1997).

Effect sizes measure the degree of association between an effect (e.g., a main effect, an interaction, a linear contrast) and the dependent variable, and are interpreted as the proportion of variance in the dependent variable that is attributable to each effect. They are of eminent importance in applied research, where they are used as additional information next to statistical significance levels. There are several measures of effect size (Kirk 1982; Tabachnick and Fidell 1989). In the R package, the most commonly used measure Eta squared (\( \eta^2_p \)), is calculated. It is defined as follows,

\[
\eta^2 = \frac{SS_{effect}}{SS_{total}},
\] (22)

where: \( SS_{effect} \) = the sums of squares for the effect of interest, \( SS_{total} \) = the total sums of squares for all effects, interactions, and errors in the regression.

For example, the effect size of factor \( \theta^p_{pq,j} \) is, \( \eta^2_{pq,j} = \frac{SS(\theta^p_{pq,j})}{SS(\theta^p_{pq,j,m})} \). Because the effect sizes are calculated at each draw of the parameters, their posterior distributions are obtained. In the package, the function \texttt{BAnova()} performs all computations discussed above and outputs a table of sums of squares and effect sizes,

\texttt{R> BAnova(res) // res is a fitted object from any of the models}

3.10. Table of P-values

The package computes Bayesian p-values for posteriors of each factor (Gill 2007), which enables significance testing. The null-hypotheses for the test concerning \( \theta^p_{pq,j,k} \) in equation (3)
$H_0: \theta_{j,k}^{pq} = 0$, versus $H_1: \theta_{j,k}^{pq} \neq 0$.  

The two-sided Bayesian P-value is obtained by first finding the one sided P-value, $\min(P(\theta_{j,k}^{pq} < 0), P(\theta_{j,k}^{pq} > 0))$ which is estimated from posterior samples as: $\min(\sum_m (I(\theta_{j,k,m}^{pq} < 0)/M, \sum_m I(\theta_{j,k,m}^{pq} > 0)/M))$. Then, the two sided P-value is

$$P_\theta(\theta_{j,k}^{pq}) = 2 \times \min(P(\theta_{j,k}^{pq} < 0), P(\theta_{j,k}^{pq} > 0)).$$  

If there are coefficients $\theta_{j,k_1}^{pq}, \theta_{j,k_2}^{pq}, \ldots, \theta_{j,k_J}^{pq}$ representing $J$ levels of a factor with more than two levels, we calculate a single P-value to represent the significance differences among all levels, as in standard ANOVA. The null-hypothesis is:

$$H_0: \theta_{j,k_1}^{pq} = \theta_{j,k_2}^{pq} = \ldots = \theta_{j,k_J}^{pq} = 0, \text{ versus } H_1: \text{some } \theta_{j,k_j}^{pq} \neq 0. $$

We compute these Bayesian p-values in this case as follows. Let $\theta_{j,k_{\min}}^{pq}$ and $\theta_{j,k_{\max}}^{pq}$ denote the coefficients with the smallest and largest posterior mean. Then the P-value is defined as

$$\min(P_\theta(\theta_{j,k_{\min}}^{pq}), P_\theta(\theta_{j,k_{\max}}^{pq})).$$

The function `table.pvalues()` in the package computes p-values for all factors and outputs a table of p-values.

R> table.pvalues(res) // res is a fitted object from any of the models

4. Applications

In this section, we provide three applications of the BANOVA package to the analysis of previously published experimental studies. The first study by Etkin and Ratner (2012) investigated how the perceived variety among products, as means to a goal, affects peoples’ motivation to pursue that goal. In this application we illustrate a between-subjects ANOVA, with dependent variables that are, respectively Normal and ordered categorical. The second study, by Ferraro, Kirmani, and Matherly (2013), examines the effects of conspicuous brand usage on consumers’ attitudes toward a brand. In this application, we illustrate hierarchical ANCOVA models, with Normal and t-distributed dependent variables. The third study by Wedel and Pieters (2014), investigates the effects of color on the rapid gist perception of advertising. In this study, we illustrate the application of a hierarchical ANOVA with both within- and between-subjects factors, and a binomial dependent variable.

4.1. Application 1: Impact of the variety among means on motivation

In this examples we illustrate the application of BANOVA package on a data from a study on goal attainment (Etkin and Ratner 2012). The study investigated how the perceived variety (high vs. low) among products, as means to a subjects’ goal, affects their motivation to pursue that goal. The hypothesis was that only when progress toward a goal is low, product variety increases motivation to pursue the goal. In the study, one hundred and five subjects were
randomly assigned to one of four conditions in a 2 (goal progress: low vs. high) by 2 (variety among means: low vs. high) between-subjects design. The final goal was a "fitness goal", and the products used were protein bars; variety was manipulated by asking subjects to think about how the products were similar (low) or different (high); goal progress was primed by asking subjects questions regarding the frequency of their recent workouts on low (0, 1,..., 5 or more) versus high (5 or less, 6, 7,..., 10) frequency scales. Subjects were asked questions regarding the similarity of protein bars as a manipulation check, and the bid they were willing to make for the bars, which are used as dependent variables in the study.

The data can be loaded by the following R command:

```R
R> data(goalstudy)
```

The structure of the data is shown below:

```R
R> head(goalstudy)
       id perceivedsim goalprogress varmeans bid
1       1            5         1 1 2 5
2       2            7         1 1 1 0
3       3            2         2 2 2 1
4       4            2         2 1 1 5
5       5            5         2 1 3
6       6            5         1 1 1 0
```

The between-subjects variables are: goalprogress, which denotes the progress toward a goal (1:low, 2: high); varmeans, which denotes the amount of variety within the means to goal attainment (1:low, 2:high); perceivedsim, which is a seven-point scale dependent variable measuring the perceived similarity of the products (1 = not at all similar, 7 = very similar); and bid which denotes the amount that subjects would be willing to pay for the products.

In the first analysis, we consider log transform of the bid amount (log(bid + 1)) as the dependent variable, assumed to follow a normal distribution. This analysis comprises a 2 (goal progress: low vs. high) x 2(variety among means: low vs. high) between-subjects hierarchical Bayesian ANOVA of the bid amount. Since the study involves a between-subjects design, the within-subjects model only includes an intercept. The function `BANOVA.normal()` is used to execute the analysis:

```R
R> goalstudy$logbid <- log(goalstudy$bid + 1)
R> app_1 <- BANOVA.Normal(logbid~1, ~goalprogress*varmeans, goalstudy, +                     goalstudy$id, burnin = 5000, sample = 1000, thin = 20)
```

The posterior means and standard deviations of the hyper parameters are reported from 1000 target samples, with a thinning factor of 20 to reduce autocorrelation, and with 5,000 samples being discarded as the burn-in period, for a total of 25,000 samples. To confirm that the chain has converged after the burn-in, the following R command outputs the Geweke's and the Heidelberg and Welch convergence diagnostics. The results are shown below,
R> conv.diag(app_1)

Geweke Diag.

|                                | Stationarity test | Convergence | p value |
|--------------------------------|-------------------|-------------|---------|
| (Intercept) : (Intercept)      | Passed            |             | 0.3696  |
| (Intercept) : goalprogress1    | Passed            |             | 0.3833  |
| (Intercept) : varmeans1        | Passed            |             | 0.6585  |
| (Intercept) : goalprogress1:varmeans1 | Passed         |             | 0.283   |

Heidelberger and Welch's Diag.

|                                | stest | start | pvalue |
|--------------------------------|-------|-------|--------|
| (Intercept) : (Intercept)      | Passed| 1     | 0.1488 |
| (Intercept) : goalprogress1    | Passed| 1     | 0.2445 |
| (Intercept) : varmeans1        | Passed| 1     | 0.073  |
| (Intercept) : goalprogress1:varmeans1 | Passed| 1     | 0.0672 |

The above result indicates that the chains converged well before the end of the burn-in. The function `trace.plot()` provides visual diagnostics of convergence, some of the results are shown in Figure 2.

Figure 2: Trace plots of (selected) posterior distributions of the parameters of the goalstudy application

The posterior means, standard deviations, 95% credible intervals and Bayesian p-values of
hyper parameters are computed as follows, and the results are shown below. Following standard conventions, we will call an effect ‘significant’ if the 95% posterior credible interval of the parameter does not cover zero.

\[ R \text{> } \text{summary(app_1)} \]

Call:
BANOVA.Normal(l1_formula = logbid ~ 1, l2_formula = ~goalprogress * varmeans, data = goalstudy, id = goalstudy$id, burnin = 5000, sample = 1000, thin = 20)

Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):
(Intercept) 0.641 (1.16%) 0.443 (0.8%) 7.584 (13.68%) 46.749 55.455

Table of p-values (Multidimensional):
(Intercept) goalprogress varmeans goalprogress:varmeans
(Intercept) 0 0.538 0.706 0.004

Table of coefficients:

|                | mean       | SD          | Quantile0.025 | Quantile0.975 | p.value | Signif.codes |
|----------------|------------|-------------|---------------|---------------|---------|--------------|
| (Intercept)    | 1.08841    | 0.09027     | 0.91787       | 1.26383       | <2e-16  | ***          |
| (Intercept)    | 0.056      | 0.09002     | -0.12835      | 0.23335       | 0.538   |              |
| (Intercept)    | 0.03147    | 0.08717     | -0.14056      | 0.20189       | 0.706   |              |
| (Intercept)    | -0.25917   | 0.08558     | -0.42512      | -0.08899      | 0.004   | **           |

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Based on the table of p-values and coefficients in the above results, the interaction between variety among means (varmeans) and goal progress (goalprogress) is significant. The table of means, produced with the command below, shows that when goal progress was low (goalprogress = 1), participants bid more for the products when perceived variety was high (varmeans = 2) versus low (varmeans = 1). On the contrary, when goal progress was high, participants bid more when perceived variety was low versus high.

\[ R \text{> } \text{table.means(app_1)} \]

Grand mean:
1.08841
2.5% 97.5%
0.91787 1.26383

Means for factors at level 2:

|                |          |          |            |            |
|----------------|----------|----------|------------|------------|
| goalprogress   | median   | 2.5%     | 97.5%      | 1.15012    | 0.88968   | 1.39588   |
|                |          |          |            | 1.02906    | 0.79407   | 1.27638   |
| varmeans       | median   | 2.5%     | 97.5%      | 1.12212    | 0.87387   | 1.35245   |
2 1.05696 0.7941 1.30155

Means for interactions at level 2:

| goalprogress | varmeans | median | 2.5%  | 97.5% |
|--------------|----------|--------|-------|-------|
| 1            | 1        | 0.92061| 0.56703| 1.23998|
| 1            | 2        | 1.37114| 0.99789| 1.74485|
| 2            | 1        | 1.3193 | 0.97855| 1.67864|
| 2            | 2        | 0.7477 | 0.406  | 1.07679|

To predict specific values of the dependent variable, the function `predict()` in R can be applied to the objects returned by `BANOVA.*()`. For example, to predict the value of the dependent variable for the 3rd subject in the data set, respectively a situation of low goal progress and a high variety among means, the following R commands can be used:

```r
R> # predict the mean for the 3rd subject (id == 3)
R> predict(app_1, goalstudy[3,])
```

```
Median  2.5%  97.5%
[1,] 0.72486 0.3996 1.08537
```

To predict the value of the dependent variable for a situation of low goal progress and a high variety among means, respectively the entire data set, the following R commands can be used (the results are not shown):

```r
R> # predict the mean corresponding to goalprogress:1 and varmeans:2
R> predict(app_1, c(0,0,1,2,0,0)) #all variables must have a value, but only
# the values of goalprogress and varmeans will be considered
R> # predict all training data
R> predict(app_1, goalstudy)
```

Since even the log-normal distribution may not describe the bid data very well, it could also be analyzed assuming a Poisson distribution for the bid amounts (there are only a few non-integer values which are rounded). The following R commands constructs the hierarchical Bayes ANOVA model and summarizes the results shown below.

```r
R> goalstudy$bid <- as.integer(goalstudy$bid + 0.5)
R> app_1a<-BANOVA.Poisson(bid~1, ~goalprogress*varmeans, goalstudy, + goalstudy$id, burnin = 5000, sample = 1000, thin = 20)
R> summary(app_1a)
```

Call:
`BANOVA.Poisson(l1_formula = bid ~ 1, l2_formula = ~goalprogress * varmeans, data = goalstudy, id = goalstudy$id, burnin = 5000, sample = 1000, thin = 20)`
Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):

|               | goalprogress | varmeans | goalprogress.varmeans | Residuals | Total  |
|---------------|--------------|----------|-----------------------|-----------|--------|
| (Intercept)   | 2.087 (1.82%)| 1.692 (1.48%)| 21.245 (18.56%) | 89.227  | 114.455 |

Table of p-values (Multidimensional):

|               | goalprogress | varmeans | goalprogress.varmeans |
|---------------|--------------|----------|-----------------------|
| (Intercept)   | 0.002        | 0.53     | 0.672                 | 0.002     |

Table of coefficients:

|               | mean         | SD         | Quantile0.025         | Quantile0.975 | p.value | Signif.codes |
|---------------|--------------|------------|-----------------------|----------------|---------|---------------|
| (Intercept)   | (Intercept)  | 0.57154    | 0.15489               | 0.87931       | 0.002   | **            |
| (Intercept)   | goalprogress | 0.08994    | 0.14916               | -0.18453      | 0.39551 | 0.53          |
| (Intercept)   | varmeans     | 0.06434    | 0.14453               | -0.21212      | 0.34755 | 0.672         |
| (Intercept)   | goalprogress1| -0.43875   | 0.14365               | -0.73404      | -0.16503| 0.002 **      |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

The next analysis is a manipulation check: the perceived similarity of the products is the dependent variable expected to be different between the levels of the varmeans factor. Since it is a seven-point scale variable, an ordered multinomial distribution is used. The 2 (goal progress: low vs. high) x 2(variety among means: low vs. high) hierarchical Bayesian ANOVA is executed using the function `BANOVA.ordMultinomial()` in the BANOVA package. Since the study involves a between-subjects design, the within-subjects model only includes an intercept. All between-subjects factors are included in the level-2 model. The analysis is done with the following commands, the results are provided below.

```r
R> app_2 <- BANOVA.ordMultinomial (perceivedsim~1, ~goalprogress*varmeans, + goalstudy, goalstudy$id, burnin = 3000, sample = 1000, thin = 5)
R> summary(app_2)
```

Call:

`BANOVA.ordMultinomial(l1_formula = perceivedsim ~ 1, l2_formula = ~goalprogress * varmeans, data = goalstudy, id = goalstudy$id, burnin = 3000, sample = 1000, thin = 5)`

Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):

|               | goalprogress | varmeans | goalprogress.varmeans | Residuals | Total  |
|---------------|--------------|----------|-----------------------|-----------|--------|
| (Intercept)   | 6.462 (1.05%)| 195.668 (31.82%)| 12.191 (1.98%) | 406.159  | 614.945 |

Table of p-values (Multidimensional):

|               | goalprogress | varmeans | goalprogress.varmeans |
|---------------|--------------|----------|-----------------------|
| (Intercept)   | 0.002        | 0.53     | 0.672                 |

Table of coefficients:

|               | mean         | SD         | Quantile0.025         | Quantile0.975 | p.value | Signif.codes |
|---------------|--------------|------------|-----------------------|----------------|---------|---------------|
| (Intercept)   | (Intercept)  | 4.66217    | 0.64226               | 3.40235       | 5.85443 | <2e-16 ***    |
| (Intercept)   | goalprogress1| -0.1665    | 0.27006               | -0.70745      | 0.39547 | 0.51          |
| (Intercept)   | varmeans1    | 1.34737    | 0.32153               | 0.71751       | 1.98923 | <2e-16 ***    |
| (Intercept)   | goalprogress1| 0.27778    | 0.28849               | -0.29295      | 0.89242 | 0.332         |

Cutpoint[2] 2.197867 0.489979 1.328096 3.230276 <NA> <NA>
Cutpoint[3] 3.576643 0.600432 2.410282 4.794221 <NA> <NA>
Cutpoint[4] 4.900657 0.670827 3.524873 6.141887 <NA> <NA>
Cutpoint[5] 6.736383 0.826927 5.026151 8.111762 <NA> <NA>
The chain converged well within the burn-in period (the convergence statistics are not shown here). The posterior means and standard deviations of the hyper parameters are reported from a total of 8,000 samples, with 3,000 being discarded as the burn-in period, and the remaining 5,000 samples thinned by a factor 5. From the Bayesian p-values, we can see that only the variety condition (intercept: varmeans) has a significant effect on the perceived similarity of the products and the table of sums of squares shows that the effect size is relatively large. The table of means is produced with the following command, and the result is provided below.

\[ R > \text{table.means(app_2)} \]

**Table of means of the response**

|                       | 2%    | 97.5% |
|-----------------------|-------|-------|
| Grand mean:           | 4.23097 |
| 2.5% 97.5%            | 3.87904 4.57188 |

**Means for factors at level 2:**

| goalprogress | median | 2.5% | 97.5% |
|--------------|--------|------|-------|
| 1            | 4.1277 | 3.60441 | 4.5929 |
| 2            | 4.32479 | 3.85958 | 4.7886 |

| varmeans    | median | 2.5% | 97.5% |
|-------------|--------|------|-------|
| 1           | 5.03269 | 4.56305 | 5.4397 |
| 2           | 3.35031 | 2.88525 | 3.87616 |

**Means for interactions at level 2:**

| goalprogress | varmeans | median | 2.5% | 97.5% |
|--------------|----------|--------|------|-------|
| 1            | 1        | 5.10153 | 4.41012 | 5.64914 |
| 1            | 2        | 3.08967 | 2.39879 | 3.80449 |
| 2            | 1        | 4.96424 | 4.37125 | 5.49808 |
| 2            | 2        | 3.64608 | 2.94386 | 4.35304 |

From the table of means of the factor varmeans, we conclude that subjects perceived the products as more similar when asked to think about how they were similar versus different, which supports the experimental manipulation and is consistent with Etkin and Ratner (2012). The function `table.means()` in the case of an ordered categorical variable also provides a more detailed table with the probabilities of each response category. For example, the table
of probabilities corresponding to response 1 (not at all similar) of the variable varmeans, is shown below (the remainder of the output for the other six response categories is not shown). From the table, we can see that subjects are more likely to provide the response value 1 when perceived variety was high, which is in line with the results above.

Table of probabilities for each category of the response

Response : 1

Grand mean:

|         | 2.5% | 97.5% |
|---------|------|-------|
| 0.00915 | 0.00284 | 0.03222 |

Means for factors at level 2:

| goalprogress | median | 2.5% | 97.5% |
|--------------|--------|------|-------|
| 1            | 0.01125 | 0.00297 | 0.04087 |
| 2            | 0.00807 | 0.00218 | 0.03081 |

| varmeans | median | 2.5% | 97.5% |
|----------|--------|------|-------|
| 1        | 0.00232 | 0.00056 | 0.01304 |
| 2        | 0.03603 | 0.01011 | 0.10057 |

Means for interactions at level 2:

| goalprogress | varmeans | median | 2.5% | 97.5% |
|--------------|----------|--------|------|-------|
| 1            | 1        | 0.00219 | 4e-04 | 0.0136 |
| 1            | 2        | 0.05444 | 0.0126 | 0.17123 |
| 2            | 1        | 0.00272 | 0.00053 | 0.01469 |
| 2            | 2        | 0.0236  | 0.0056  | 0.08551 |

To predict means corresponding to the first two data points in goalstudy, the predict() command is used, which outputs the probabilities for each category of each data point.

R> predict(app_2, goalstudy[1:2,])

| Sample number | Response | Median | 2.5%    | 97.5%    |
|---------------|----------|--------|---------|----------|
| [1,]          | 1        | 0.05626| 0.01281 | 0.18767  |
| [2,]          | 1        | 0.27994| 0.10976 | 0.50518  |
| [3,]          | 1        | 0.30298| 0.16622 | 0.46136  |
| [4,]          | 1        | 0.20120| 0.08078 | 0.37816  |
| [5,]          | 1        | 0.09312| 0.02778 | 0.24850  |
| [6,]          | 1        | 0.01735| 0.00352 | 0.07083  |
| [7,]          | 1        | 0.00267| 0.00045 | 0.02066  |
The JAGS code for the above model generated by the program can be easily retrieved by the following R command.

```R
R> cat(app_2$JAGSmodel)
```

```r
model{
  for (i in 1:n){
    y[i] ~ dcat(P[i,])
    P[i,1] <- 1 - Q[i,1]
    for (i.cut in 2: n.cut) {
      P[i,i.cut] <- Q[i,i.cut-1] - Q[i,i.cut]
    }
    P[i,n.cut+1] <- Q[i,n.cut]
    for (i.cut in 1:n.cut){
      logit(Q[i,i.cut]) <- z[i,i.cut]
      z[i,i.cut] <- beta1[id[i]]*X[i,1]-cutp[i.cut]*(1-equals(i.cut,1))
    }
  }
  for (i in 1:M){
    beta1[i]~dnorm(mu.beta1[i],tau.beta1)
    mu.beta1[i]<- beta1_1*Z[i,1]+beta1_2*Z[i,2]+beta1_3*Z[i,3]+beta1_4*Z[i,4]
  }
  tau.beta1<-pow(sigma.beta1,-2)
  sigma.beta1~dgamma(1,1)
  beta1_1~dnorm(0,0.0001)
  beta1_2~dnorm(0,0.0001)
  beta1_3~dnorm(0,0.0001)
  beta1_4~dnorm(0,0.0001)
  for (i.cut in 1: n.cut) {
    cutp0[i.cut] ~ dunif(0,10)
  }
  cutp[1:n.cut] <- sort(cutp0)
}
```

Note that, for the convenience of generation of the JAGS code, the program uses a uniform naming scheme for all level 1 and level 2 parameters which are different from the names in the original data.
4.2. Application 2: Conspicuous brand usage

We next illustrate the BANOVA package on data from a study that examines consumers’ attitudes toward a brand after seeing another consumer conspicuously using it (Ferraro et al. 2013). Conspicuous brand usage occurs when a consumer draws attention to a brand she uses by flaunting (Ferraro et al. 2013). One hundred fifty-four subjects from an online panel participated in the study. Conspicuousness was manipulated as a between-subjects factor, by exposing subjects to a forty-five seconds video in which the conspicuous usage of the brand (Apple ipad) was manipulated (low vs high conspicuousness). In addition, the so called self-brand connection was measured: this refers to the extent to which a consumers’ own self-concept matches the image she has of a certain brand. Brand attitude was calculated as the average of three seven-point scale questions (dislike/like, unfavorable/favorable and bad/good). The relation between conspicuousness, self-brand connection, and brand attitudes was investigated. The analysis aims to test the hypothesis that there are negative effects of conspicuous brand usage on the attitudes toward the brand, only for subjects that have a low self-brand connection.

This example illustrates a hierarchical Bayesian ANCOVA model. Brand attitude is considered as a continuous dependent variable, assumed to follow a Normal distribution. Data of this study can be loaded by the following R command:

R> data(ipadstudy)

The data is displayed in the long format including only responses and between-subjects variables.

R> head(ipadstudy)

| id | attitude | owner | age | gender | conspic | selfbrand | apple_dl |
|----|----------|-------|-----|--------|---------|------------|---------|
| 1  | 1        | 0     | 19  | 0      | 0       | -2.3042672 | 3       |
| 2  | 2        | 0     | 33  | 0      | 1       | 1.6957328  | 6       |
| 3  | 3        | 0     | 25  | 1      | 0       | -0.1614100 | 6       |
| 4  | 4        | 0     | 41  | 0      | 1       | -0.4471243 | 5       |
| 5  | 5        | 1     | 38  | 1      | 1       | 0.2671614  | 6       |
| 6  | 6        | 1     | 33  | 1      | 0       | 0.6957328  | 4       |

The between-subjects variables are: selfbrand, which is a numerical variable (mean centered) representing self-brand connection; and conspic, which is a two-level factor corresponding to the two levels of conspicuousness (0: low, 1: high). There are a number of control variables: owner denotes whether participants owned the product (1: yes, 0: no); age in years; and gender (1: female, 0: male). Finally id is the identification number of the subjects.

The Bayes ANOVA uses as a dependent variable the attitude toward the brand, measured by averaging answers on three seven-point scales. It can be executed using the function BANOVA.Normal() in the BANOVA package. Since it is a between-subjects design, the within-subjects model only includes an intercept. The between-subjects covariates include owner, age, gender, selfbrand and the interaction between conspic and selfbrand. The two-level factor conspic is effects coded and the one-way between-subjects ANCOVA is specified as follows.
R> # mean center covariates
R> ipadstudy$age <- ipadstudy$age - mean(ipadstudy$age)
R> ipadstudy$owner <- ipadstudy$owner - mean(ipadstudy$owner)
R> ipadstudy$gender <- ipadstudy$gender - mean(ipadstudy$gender)
R> app_3 <- BANOVA.Normal(attitude~1, ~owner + age + gender + selfbrand*conspic,
+  ipadstudy, ipadstudy$id, burnin = 5000, sample = 1000, thin = 10 )

The posterior means and standard deviations of the hyper parameters are reported from a total
15,000 samples, with 5,000 being discarded as the burn-in period, and the remainder thinned
by a factor 10. The chain converged well within the burn-in period (the convergence statistics
are not shown here). The table of sums-of-squares, effect sizes and Bayesian p-values, as well
as the posterior means, standard deviations, 95% credible intervals, and Bayesian p-values of
hyper parameters are computed as follows, the results are shown below.

R> summary(app_3)

Call:
BANOVA.Normal(l1_formula = attitude ~ 1, l2_formula = ~owner +
  age + gender + selfbrand * conspic, data = ipadstudy, id = ipadstudy$id,
  burnin = 5000, sample = 1000, thin = 10)

Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):

|           | owner | age | gender | selfbrand | conspic |
|-----------|-------|-----|--------|-----------|---------|
| (Intercept) | 10.922 (3.44%) | 0.637 (0.2%) | 0.659 (0.21%) | 129.164 (40.73%) | 8.982 (2.83%) |
| selfbrand.conspic | Residuals | Total |
| 6.98 (2.2%) | 86.923 | 317.133 |

Table of p-values (Multidimensional):

| (Intercept) | owner | age | gender | selfbrand | conspic | selfbrand:conspic |
|-------------|-------|-----|--------|-----------|---------|-------------------|
| (Intercept) | 0.006 | 0.888 | 0.846 | 0 | 0.008 | 0.028 |

Table of coefficients:

| | mean | SD | Quantile0.025 | Quantile0.975 | p.value | Signif.codes |
| (Intercept) | : (Intercept) | 5.27773 | 0.0946 | 5.10565 | 5.47032 | <2e-16 | *** |
| (Intercept) | : owner | 0.60823 | 0.22316 | 0.18081 | 1.06966 | 0.006 | ** |
| (Intercept) | : age | -0.00117 | 0.00975 | -0.02040 | 0.01868 | 0.888 | |
| (Intercept) | : gender | -0.0325 | 0.18076 | -0.4091 | 0.31607 | 0.846 | |
| (Intercept) | : selfbrand | 0.62535 | 0.06329 | 0.49116 | 0.74733 | <2e-16 | *** |
| (Intercept) | : conspic1 | 0.23208 | 0.09137 | 0.06009 | 0.41803 | 0.008 | ** |
| (Intercept) | : selfbrand:conspic1 | -0.12997 | 0.05908 | -0.24961 | -0.02326 | 0.028 | * |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Based on the above estimates, conspicuousness, self-brand connection and the interaction:
conspicuousness x self-brand connection, significantly affect the attitude towards the brand,
consistent with Ferraro et al. (2013). Note that since we applied ANCOVA, a type III analysis
of variance is used, so that the sum of squares (and effect size) for each effect is computed
conditional upon all other effects and thus they do not add up to the total sum of squares.

In this application, the distribution of the dependent variable is continuous, but it may not
be Normal. The function BANOVA.T() can be applied to construct an ANOVA model in which
the response variable is assumed to follow a student’s t distribution. This permits (weakly)
robust inference (Bernardo and Giron 1992), as it allows for fatter tails and outliers in the data. The results, shown below, are similar to those in the results above which supports their robustness to distributional assumptions, and are not discussed here.

\[
R> app_4 \leftarrow \text{BANOVA.T}(\text{attitude} \sim 1, \text{~owner + age + gender + selfbrand*conspic, + ipadstudy, ipadstudy$id, burnin = 3000, sample = 1000, thin = 5 })
\]

\[
R> \text{summary(app}_4\text{)}
\]

Call:
BANOVA.T(l1_formula = attitude ~ 1, l2_formula = ~owner + age +
gender + selfbrand * conspic, data = ipadstudy, id = ipadstudy$id,
burnin = 5000, sample = 1000, thin = 10)

Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):
|           | owner | age | gender | selfbrand | conspic |
|-----------|-------|-----|--------|------------|---------|
| (Intercept) | 10.939 | 0.34 | 0.354  | 126.2      | 9.01    |
| selfbrand.conspic Residuals Total | 6.251 | 135.715 | 361.517 |

Table of p-values (Multidimensional):
| (Intercept) | owner | age | gender | selfbrand | conspic | selfbrand:conspic |
|-------------|-------|-----|--------|------------|---------|-------------------|
| (Intercept) | 0     | 0.942 | 0.94  | 0          | 0.01    | 0.028             |

Table of coefficients:

| (Intercept) : (Intercept) | 5.29064 | 0.08997 | 5.10989 | 5.4735 | <2e-16 | *** |
| (Intercept) : owner      | 0.61526 | 0.21369 | 0.21623 | 1.0569 | <2e-16 | *** |
| (Intercept) : age        | -0.00064 | 0.00956 | -0.01811 | 0.01893 | 0.942  |
| (Intercept) : gender     | -0.01146 | 0.18485 | -0.36935 | 0.3533 | 0.94   |
| (Intercept) : selfbrand  | 0.62077 | 0.06185 | 0.49873 | 0.73824 | <2e-16 | *** |
| (Intercept) : conspic1   | 0.23785 | 0.09262 | 0.0523  | 0.42045 | 0.01   | **  |
| (Intercept) : selfbrand:conspic1 | -0.12541 | 0.05823 | -0.23701 | -0.00892 | 0.028  | *   |

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

4.3. Application 3: Gist perception of advertising

We finally illustrate the application of the BANOVA package in a study into the influence of color on advertising gist perception, which is the very rapid identification of ads during brief exposures. Specifically, we analyze the effect of color on the perception of the gist of advertising when the advertising exposure is blurred (Wedel and Pieters 2014). In the study, one hundred and sixteen subjects were randomly assigned to one condition of a 5 (blur: normal, low, medium, high, very high) x 2 (color: full color, grayscale) between-participants, x 2 (image: typical ads, atypical ads) within-participants, mixed design. Participants were exposed to 40 images, 32 full-page ads and 8 editorial pages. There were 8 ads per product category, with 4 typical and 4 atypical ones. Blur was manipulated by processing the advertising images with a Gaussian blur filter of different radius. Subjects were asked to identify, after being flashed an image for 100msec., whether the image was an ad or not.

The data included in the package can be loaded into R using the data() function, i.e., using the following R code:
R> data(colorad)

The structure of colorad is shown below using the head() function. It is in long format including both within- subjects and between- subjects variables.

R> head(colorad)

    id typic  y blurfac  color  blur
  1     1     0     8       2     1 3.6889
  2     1     1     6       2     1 3.6889
  3     2     0    12       4     0 4.7875
  4     2     1     6       4     0 4.7875
  5     3     0    11       2     0 3.6889
  6     3     1     9       2     0 3.6889

Here, the within-subject variable typic is a factor with 2 levels '0' (typical ads) and '1'(atypical ads); between-subject variables are: blur, a numerical variable representing the blur of the image (the log-radius of a Gaussian blur filter used to produce the images), blurfac, a factor variable with the five levels of blur, and color, a factor representing the color of the ads with 2 levels '0'(full color) and '1'(grayscale). id is the subject identification number. The dependent variable is the number of times ads were correctly identified as an ad, out of the 16 ads, for each subject for each level of typic.

We are interested in the effects of within- and between- subject factors typic, and color, and the variable blur, as well as their interactions. The factor typic varies within individuals; the factors blur, color and blur x color interaction vary between individuals.

The analysis of this experiment is executed with the function BANOVA.Bin() in the BANOVA package. First, the continuous covariate blur is mean centered. The R code to implement the analysis is shown below.

R> data(colorad)
R> # mean center Blur for effect coding
R> colorad$blur <- colorad$blur - mean(colorad$blur)
R> app_5 <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, +                       as.integer(16), burnin = 3000, sample = 2000, thin = 5)

The posterior means and standard deviations of the hyper parameters are reported from a total of 13,000 samples, with 3,000 being discarded as the burn-in period, and the remainder thinned by a factor 5 to reduce autocorrelation. To confirm that the chain has converged after the burn-in, the conv.diag(app_5) command outputs the convergence diagnostics and trace.plot(app_5) provides visual diagnostics of convergence (the results are not shown here, but indicates that the chains converged well before the end of the burn-in).

The posterior means, standard deviations, 95% credible intervals and Bayesian p-values of hyper parameters are computed with the following command, and the results are shown below.
\textbf{BANOVA}: An \textit{R} Package for Hierarchical Bayesian ANOVA

\texttt{R> summary(app_5)}

Call:
\texttt{BANOVA.Bin(l1_formula = y \sim typic, l2_formula = \sim color + blur,}
\texttt{ data = colorad, id = colorad$id, num_trails = as.integer(16),}
\texttt{ burnin = 3000, sample = 2000, thin = 5)}

Table of sum of squares and effect sizes (Bayesian ANOVA/ANCOVA):
\begin{verbatim}
|            | color | blur | color:blur | Residuals | Total  |
|------------|-------|------|------------|-----------|--------|
| (Intercept)| 4.691 | 26.344 | 1.679 | 131.733 | 164.733 |
| typic1     | 1.55  | 5.651 | 1.403 | 6.307 | 14.965  |
\end{verbatim}

Table of p-values (Multidimensional):
\begin{verbatim}
| (Intercept) | color | blur | color:blur |
|-------------|-------|------|------------|
|             | 0     | 0.011 | 0.166      |
\end{verbatim}

Table of coefficients:
\begin{verbatim}
| (Intercept) : (Intercept) | 0.49476 | 0.05543 | 0.38737 | 0.60236 | <2e-16 *** |
| (Intercept) : color1 | 0.13878 | 0.05526 | 0.03331 | 0.24904 | 0.011 *   |
| (Intercept) : blur | -0.17061 | 0.02893 | -0.22741 | -0.11583 | <2e-16 *** |
| typic1 : (Intercept) | 0.40207 | 0.02917 | -0.01769 | 0.09815 | 0.166     |
| typic1 : color1 | 0.07738 | 0.02668 | 0.02536 | 0.12885 | 0.003 **  |
| typic1 : blur | -0.07775 | 0.01485 | -0.10618 | -0.04917 | <2e-16 *** |
| typic1 : color:blur | 0.03697 | 0.01441 | 0.00895 | 0.06458 | 0.011 *   |
\end{verbatim}

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Based on the above estimates, ad identification is significantly influenced by ad typicality (typic): typical ads are identified more accurately as ads as compared to less typical ones. The accuracy of ad identification is also affected by the degree of blur. The three-factor interaction (blur x color x typic) is also significant, which reveals that color protects the identification of typical ads against blur (Wedel and Pieters 2014).

These results are based on the ANCOVA model with blur as a continuous covariate. To further understand the effects of blur, we can use the discrete variable blur (blurfac) in a two-way ANOVA at the between-subject level (and the factor typic again within-subjects), using the following command:

\texttt{R> app_6 <- BANOVA.Bin(y \sim typic, \sim color*blurfac, colorad, colorad$id,}
\texttt{ + as.integer(16), burnin = 20000, sample = 3000, thin = 5)}

Since the above model involves more parameters, to ensure all parameters in the models converge, a larger number of burn-in and target samples are used: a total of 35,000. The table of sums of squares, effect sizes and p-values, as well as the posterior means, standard deviations, 95\% credible intervals and Bayesian p-values of the parameters are produced with the function \texttt{summary()}, the results are presented below.

\texttt{R> summary(app_6)}
We first inspect the tables of sums of squares, effect sizes and p-values. In this table, the columns denote between-subject factors and the rows denote the within-subject factors. The values in the table present the sum of squares and effect sizes of the effects of these between-subject factors on the within-subjects factors. Again, the accuracy of ad identification is affected by blur, and to a lesser extent by color. From the tables of p-values, ad typicality (the value corresponding to the row name ‘typic’ and column name ‘(Intercept)’) and the degree of blur (the value corresponding to the row name ‘(Intercept)’ and column name ‘blurfac’) are again highly significant. There is also support for the main effect of color. The three-factor interaction (blurfac x color x typic) is also significant, which again shows that color protects the identification of typical ads against blur (Wedel and Pieters 2014). The conclusions from the table of estimates are similar to those from the results of the previous model, but this table allows us to inspect the effects of each level of blur, and the interactive effects with color and typicality.
Through the tables of means for all factors and their interactions, we can inspect these effects in more detail. These are produced with the function `table.means()`. The results are provided below:

```r
R> table.means(app_6)
```

**Grand mean:**
0.6206999

| 2.5%         | 97.5%         |
|--------------|--------------|
| 0.5954085    | 0.6465321    |

**Means for factors at level 1:**

| typic | median | 2.5% | 97.5% |
|-------|--------|------|-------|
| 0     | 0.69313| 0.66484| 0.7188 |
| 1     | 0.54284| 0.51437| 0.5729 |

**Means for factors at level 2:**

| color | median | 2.5% | 97.5% |
|-------|--------|------|-------|
| 0     | 0.65594| 0.61981| 0.69103|
| 1     | 0.58403| 0.54472| 0.62252|

| blurfac | median | 2.5% | 97.5% |
|---------|--------|------|-------|
| 1       | 0.74044| 0.69921| 0.78352|
| 2       | 0.63263| 0.58918| 0.69183|
| 3       | 0.6244 | 0.56966| 0.67767|
| 4       | 0.60808| 0.55263| 0.66323|
| 5       | 0.48113| 0.42043| 0.53902|

**Means for interactions between level 1 and level 2 factors:**

| typic | color | median | 2.5% | 97.5% |
|-------|-------|--------|------|-------|
| 0     | 0     | 0.7451 | 0.70637| 0.77987|
| 0     | 1     | 0.63533| 0.59554| 0.67466|
| 1     | 0     | 0.55473| 0.51075| 0.59889|
| 1     | 1     | 0.53021| 0.48991| 0.5719 |

| typic | blurfac | median | 2.5% | 97.5% |
|-------|---------|--------|------|-------|
| 0     | 0       | 0.82398| 0.77602| 0.86012|
| 0     | 1       | 0.77421| 0.71804| 0.8217 |
| 0     | 2       | 0.68515| 0.62832| 0.737 |
| 0     | 3       | 0.64344| 0.58566| 0.69783|
| 0     | 4       | 0.48101| 0.41132| 0.54899|
| 1     | 1       | 0.63501| 0.57172| 0.69241|
| 1     | 2       | 0.46266| 0.39622| 0.53239|
| 1     | 3       | 0.55919| 0.4989 | 0.61972|
| 1     | 4       | 0.57217| 0.51064| 0.63336|
| 1     | 5       | 0.48065| 0.41256| 0.54653|

**Means for interactions at level 2:**

| color | blurfac | median | 2.5% | 97.5% |
|-------|---------|--------|------|-------|
| 0     | 0       | 0.73962| 0.66789| 0.79778|
| 0     | 2       | 0.69298| 0.60684| 0.76707|
| 0     | 3       | 0.6415 | 0.56029| 0.71427|
Means for interactions between level 2 interactions and level 1 factors:

| typical | color | blurfac | median 2.5% | 97.5% |
|---------|-------|---------|-------------|-------|
| 0       | 0     | 1       | 0.82222 0.74993 0.87523 |
| 0       | 0     | 2       | 0.83973 0.76968 0.89083 |
| 0       | 0     | 3       | 0.71491 0.6369 0.78504 |
| 0       | 0     | 4       | 0.7225 0.64393 0.78792 |
| 0       | 0     | 5       | 0.57343 0.47974 0.66626 |
| 0       | 1     | 1       | 0.82414 0.76026 0.87769 |
| 0       | 1     | 2       | 0.69055 0.60125 0.76846 |
| 0       | 1     | 3       | 0.65432 0.56975 0.73316 |
| 0       | 1     | 4       | 0.5565 0.47031 0.64192 |
| 0       | 1     | 5       | 0.39874 0.29817 0.48162 |
| 1       | 0     | 1       | 0.63516 0.54555 0.71833 |
| 1       | 0     | 2       | 0.49421 0.39296 0.59681 |
| 1       | 0     | 3       | 0.56118 0.47048 0.64477 |
| 1       | 0     | 4       | 0.56344 0.47597 0.65078 |
| 1       | 0     | 5       | 0.5185 0.42485 0.61034 |
| 1       | 1     | 1       | 0.63486 0.54185 0.71608 |
| 1       | 1     | 2       | 0.43197 0.3438 0.52187 |
| 1       | 1     | 3       | 0.55946 0.46972 0.64154 |
| 1       | 1     | 4       | 0.58095 0.49503 0.66564 |
| 1       | 1     | 5       | 0.44181 0.35091 0.54084 |

From the above tables, we can see that typical color ads (typic = 0, color = 0) are always more accurately identified than atypical color ads (typic = 1, color = 0). Typical grayscale ads (typic = 0, color = 1, blur = 1,...,5), however, are only more accurately identified than atypical grayscale ads (typic = 1, color = 1, blur = 1,...,5) when there is no blur, or a low level of blur (Wedel and Pieters 2014).

5. Discussion

This paper has presented the R package BANOVA, which can be used to analyze experimental data with a wide variety of hierarchical Bayesian ANOVA models. The response variable can be normal, student’s t, binomial, multinomial or Poisson and the between-subject model follows the traditional ANOVA or ANCOVA structure that allows the estimation of sums of squares and effect sizes of each experimental factor. In the behavioral sciences there is an abundance of research that lends itself to the application of these types of analyses. We hope that the availability of user-friendly software in the form of the BANOVA package will simulate the analysis of these studies under more reasonable assumptions on the distribution of the data and its hierarchical structure. This is expected to reduce both the type I and type II errors made in this research.

There are a number of other R packages that can be used to fit hierarchical Bayes mod-
els, including **BACC** (Geweke 1999), **bayesm** (Rossi and McCulloch 2006) and **MCMCpack** (Martin, Quinn, and Park 2011) as well as the **WinBUGS**, **OpenBUGS** and **JAGS** software. However, all these either require the user to be familiar with Bayesian statistical modeling and **BUGS** programming, while some are not suitable for ANOVA analysis of experimental data. The motivation for the development of the **BANOVA** package is therefore to overcome these limitations and to offer user-friendly routines for the applied researcher. These were illustrated in three applications: (1) models can be run by only a single function call, (2) only a small number of settings are required to run each model, (3) outputs are included in summary tables, such as the table of effect sizes, table of p-values and tables of means, which are easy to interpret for the applied researcher.

The package utilizes the **JAGS** software, which offers many advantages from a modeling perspective, while in addition the **JAGS** code is produced as a by-product of the analysis and available for the applied researcher to inspect and modify. However, a disadvantage of this choice is that this decreases the computation efficiency. Future developments could focus on reprogramming the MCMC code in languages such as **C++** or **Java**. In addition, the package can be extended with other models, for example using different distributions of the dependent variable.

**Acknowledgments**

The authors are grateful to Rebecca Ratner, Jordan Etkin, Rosie Ferraro, Amna Kirmani, Ted Matherly and Rik Pieters for making their data available for this research, and to Peter Lenk for his suggestions on the computation of Bayesian P-values.

**References**

Bartlett M (1947). “The use of Transformations.” *Biometrics, 3*, 39–52.

Bernardo JM, Giron FJ (1992). Robust Sequential Prediction from Nonrandom Samples: The Election Night Forecasting Case. In Bayesian Statistics 4, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), Oxford University Press.

Bernardo J SA (2000). *Bayesian Theory*. John Wiley and Sons, West Sussex, England.

Box G, Cox D (1964). “An Analysis of Transformation.” *Journal of the Royal Statistical Society, B*(26), 211–252.

Breslow N, Clayton D (1993). “Approximate Inference in Generalized Linear Mixed Models.” *Journal of the American Statistical Association, 88*(421), 9–25.

Cardinal RN, Aitken MR (2005). *ANOVA for the Behavioral Sciences Researcher*. Psychology Press.

Cochran W (1940). “The Analysis of Variances When Experimental Errors Follow the Poisson or Binomial Laws.” *The Annals of Mathematical Statistics, 11*, 335–347.
Conover WJ, Iman RL (1976). “On Some Alternative Procedures using Ranks for the Analysis of Experimental Designs.” *Communications in Statistics*, **A5**, 1349–1368.

Denwood M (2013). “*runjags*: An R Package Providing Interface Utilities, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS.” *Journal of Statistical Software* (In Review).

Draper N, Hunter W (1969). “Transformations. Some Examples Revisited.” *Technometrics*, **11**, 23–40.

Draper N, Smith H (1998). *Applied Regression Analysis*. 3 edition. John Wiley.

Etkin J, Ratner RK (2012). “The Dynamic Impact of Variety among Means on Motivation.” *Journal of Consumer Research*, **38**(6), 1076–1092.

Ferraro R, Kirmani A, Matherly T (2013). “Look at Me! Look at Me! Conspicuous Brand Usage, Self-Brand Connection, and Dilution.” *Journal of Marketing Research*, **50**, 477–488.

Fisher RA (1921). “On the ”Probable Error” of a Coefficient of Correlation Deduced from a Small Sample.” *Metron*, **1**, 3–32.

Fisher RA (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fox J (1997). *Applied Regression Analysis, Linear Models and Related Methods*. Sage.

Gelman A (2005). “Analysis of Variance: Why It is More Important Than Ever?” *The Annals of Statistics*, **1**(33), 43–47.

Gelman A, Carlin JB, Stern HS, Dunson DB (2013). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Wiley, New York.

Geweke J (1991). *Evaluating the Accuracy of Sampling-based Approaches to Calculating Posterior Moments*. In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Geweke J (1999). “Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication.” *Econometric Reviews*, **18**, 1–126.

Gill J (2007). *Bayesian Methods for the Social and Behavioral Sciences*. 2 edition. Chapman and Hall.

Hartley H, Rao J (1967). “Maximum Likelihood Estimation for the Mixed Analysis of Variance Model.” *Biometrika*, **54**, 93–108.

Heidelberger P, Welch P (1983). “Simulation Run Length Control in the Presence of an Initial Transient.” *Ops Res.*, **31**, 1109–1144.

Ito P (1980). *Robustness of ANOVA and MANOVA test procedures*. In Handbook of Statistics, Vol. 1, (P. R. Krishnainth, ed.), Amsterdam, Holand.

Kirk RE (1982). *Experimental design: Procedures for the Behavioral Sciences*. 2 edition. Brooks/Cole, Belmont, CA.
Kruschke J (2013). “Bayesian Estimation Supersedes the t-test.” *Journal of Experimental Psychology: General*, **142**(2), 573–603.

Lindley D, Smith A (1972). “Bayes Estimates for the Linear Model.” *Journal of the Royal Statistical Society*, **B34**(1), 1–41.

Longford N (1987). “A Fast Scoring Algorithm for Maximum Likelihood Stimation in Unbalanced Models with Nested Random Effects.” *Biometrika*, **74**, 817–827.

Lord F, Novick M (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley: Reading, MA.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000). “Winbugs-a Bayesian Modeling Framework: Concepts, Structure, and Extensibility.” *Statistics and Computing*, **10**, 325–337.

Martin AD, Quinn KM, Park JH (2011). “MCMCpack: Markov Chain Monte Carlo in R.” *Journal of Statistical Software*, **42**(9), 22. URL http://www.jstatsoft.org/v42/i09/.

McCullagh P, Nelder J (1989). *Generalized Linear Models*. 2 edition. Chapman and Hall, New York, NY.

Nunnaly J (1967). *Psychometric Theory*. McGraw Hill, New York, NY.

Overall JE, Spiegel DK, Cohen J (1975). “Equivalence of Orthogonal and Nonorthogonal Analysis of Variance.” *Psychological Bulletin*, **82**(2), 182–186.

Plummer M (2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.”

Plummer M, Best N, Cowles K, Vines K (2006). “CODA: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, **6**, 7–11.

Press S (2003). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Wiley, New York.

Raudenbusch S (1988). “Eduational Applications of Hierarchical Linear Models: A review.” *Journal of Educational Statistics*, **13**, 85–116.

Raudenbush SW (1999). *Hierarchical Models*. In S. Kotz (Ed.), Encyclopedia of Statistical Sciences, Update Volume 3 (pp. 318-323). John Wiley and Sons, Inc, New York.

Rossi P, McCulloch R (2006). *bayesm*: Bayesian Inference for Marketing/Micro- Econometrics. *R* package version 2.0-8. URL http://CRAN.R-project.org/package=bayesm.

Rouder JN, Speckman P, Sun D, Morey R, Iverson G (2009). “Bayesian t Tests for Accepting and Rejecting the Null Hypothesis.” *Psychoneurosis Bulletin and Review*, **16**, 225–237.

Savage L (1954). *The Foundations of Statistics*. Wiley, New York (reprinted by Dover, New York, 1972).

Scheffé H (1957). *The Analysis of Variance*. Wiley, New York.

Tabachnick BG, Fidell LS (1989). *Using Multivariate Statistics*. 2 edition. Harper and Row, New York.
Tan W (1982). “Sampling Distributions and Robustness of t, f and Variance-ratio in Two Samples and ANOVA Models with Respect to Departure from Normality.” *Communications in Statistics Theory and Methods, 11*, 2485–2511.

Thorndike R (1971). *Educational Measurement*. American Council on Education, Washington, D.C.

Tiku M (1971). “Power Function of the F-test under Non-normal Situations.” *Journal of the American Statistical Association, 66*, 913–915.

Wedel M, Pieters FP (2014). “The Role of Color Schemes when Advertising Exposures are Brief and Blurred.” Working paper, Robert H. Smith School of Business, University of Maryland.

**Affiliation:**

Chen Dong  
Department of Mathematics  
Mathematics Building  
University of Maryland  
College Park, MD 20742-4015, USA  
E-mail: cdong@math.umd.edu

Michel Wedel  
Department of Marketing  
Robert H. Smith School of Business  
3303 Van Munching Hall  
University of Maryland  
College Park, MD 20742-1815, USA  
E-mail: mwedel@rhsmith.umd.edu  
URL: http://www.rhsmith.umd.edu/directory/michel-wedel