

ABSTRACT

Title of dissertation: Hierarchical Bayes Analysis of Behavioral Experiments

Chen Dong, Doctor of Philosophy, 2014

Dissertation directed by: Professor Michel Wedel

In this dissertation, 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** (Just Another Gibbs Sampler) which is very similar to the famous software **WinBUGS**. 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. We also develop generalized Bayesian mediation models for analysis of mediation effects. By using Bayesian analysis, inference is

straightforward and exact, which makes it appealing for experimental studies with small samples. The Bayesian approach is also conceptually simpler for any model with a complicated structure, especially for multilevel mediation analysis. Analysis of several data sets are used to illustrate the proposed methods.

Hierarchical Bayes Analysis of Behavioral Experiments

by

Chen Dong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Michel Wedel, Chair/Advisor
Professor Yuan Liao
Professor Kasso A. Okoudjou
Professor Subramanian Raghavan
Professor Paul J. Smith

© Copyright by
Chen Dong
2014

Acknowledgments

I owe my gratitude to all the people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Michel Wedel. Without his patient guidance, teaching, insightful ideas and long hours of work, this dissertation could not have been completed. I can not adequately express my thanks for his help and interest in seeing me obtain not only my degree, but succeed in all my endeavors. Since the moment that he replied my first email sent to him for advising, he gave me great help and invaluable advice in not only my research but also my life. It is from him that I learned how to build the first model and conduct data analysis for marketing purposes independently. It is because of him that I successfully completed the work that not only serve for our own research but also available for all other researchers with more or less statistical knowledge. It is because of him that I can become a better and stronger person in both academia and life. It has been a great pleasure to work with and learn from such an extraordinary individual.

I am also 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. My sincere thanks also goes to graduate students in the marketing department including Ajay Abraham, James Kim and Tom Kim. I appreciate their help and advice in my research projects.

I would like to thank Professor Paul J. Smith, Professor Subramanian Raghavan, Professor Yuan Liao, Professor Kasso A. Okoudjou for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript. Further, I would like to thank Professor Konstantina Trivisa, Ms. Alverda McCoy and Ms. Regalado Celeste for their advice and help on my graduate life here.

The most important thanks goes to my parents. Their love, endless support and encouragement always help me fight against difficulties, impossible odds and make miracles one after another in the way of chasing my dreams. Words cannot express the gratitude I owe them.

It is impossible to remember all, and I apologize to those I've inadvertently left out. Thank you all!

Table of Contents

List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Brief Review of Bayesian Statistical Methods	5
1.2.1 Bayes' Theorem	5
1.2.2 Bayes Estimates and Markov Chain Monte Carlo (MCMC)	6
1.3 Overview	9
2 BANOVA : An R Package for Hierarchical Bayesian ANOVA	11
2.1 Introduction	11
2.2 Bayesian ANOVA Models	12
2.3 BANOVA R Package	17
2.3.1 Obtaining the Software	17
2.3.2 Functionality	18
2.3.3 Data Input	19
2.3.4 Estimation of the Coefficients	21
2.3.5 Starting Values and Burn-in Period	22
2.3.6 JAGS Code	23
2.3.7 Convergence Diagnostics	23
2.3.8 Tables of Means	25
2.3.9 Table of Sums of Squares and Effect Sizes	27
2.3.10 Table of P-values	29
2.4 Applications	30
2.4.1 Application 1: Impact of the Variety among Means on Motivation	30
2.4.2 Application 2: Conspicuous Brand Usage	42
2.4.3 Application 3: Gist Perception of Advertising	47
2.5 Conclusion	56

3	Bayesian Mediation Analysis	58
3.1	Introduction	58
3.2	Single Level Mediation Models	62
3.2.1	Conventional Mediation Analysis of Single-Level Models	62
3.2.2	Bayesian Mediation Analysis of Single-Level Models	64
3.3	Bayesian Mediation Analysis of Multilevel Models	67
3.4	Applications	76
3.4.1	Application 1: Single-Level Mediation	76
3.4.2	Application 2: Multilevel Mediation	80
3.5	Conclusion	82
4	Future Work	85
	Appendix A JAGS Code for Bayesian Mediation Analysis	87
A.1	JAGS Code for Single-level Mediation	87
A.2	JAGS Code for Multilevel Mediation	89
	Appendix B BANOVA User Manual	93
	Appendix C BANOVA Web Application	123
	Bibliography	129

List of Tables

2.1	Sample of the <code>goalstudy</code> data	31
2.2	Sample of the <code>ipadstudy</code> data	43
2.3	Sample of the <code>colorad</code> data	48
3.1	Descriptive statistics of housing data	77
3.2	Conventional mediation analysis results of housing data	77
3.3	Bayesian mediation analysis results of housing data	78
3.4	Comparison of conventional and Bayesian mediation analysis of housing data with dichotomous variables	80
3.5	Bayesian multilevel mediation analysis of housing data	82

List of Figures

2.1	Example of the format of input data	20
2.2	Trace plots of (selected) posterior distributions of the parameters of the <code>goalstudy</code> application	33
2.3	JAGS code for the <code>goalstudy</code> application with <code>perceivedsim</code> as the dependent variable	42
3.1	Mediation diagram	59
3.2	Density plot of the posterior distribution of the mediated effect and corresponding normal quantile-quantile plot	79
3.3	Density plot of the posterior distribution of the mediated effect and corresponding normal quantile-quantile plot	81
3.4	Trace and density plot of the posterior distribution of the mediated effect $z_{a \times b}$ in between-subject level multilevel mediation model	83
C.1	Data input	124
C.2	Select subject ID	124
C.3	Select a dependent variable and a model	125
C.4	Select between-subject variables	126
C.5	Select interactions among between-subject variables	126
C.6	Select within-subject variables	127
C.7	Parameters setting	128
C.8	Results display	128

Chapter 1: Introduction

1.1 Background and Motivation

Because of its computational attractiveness and the ease of interpretation of its statistical tests and the corresponding tables of means, Analysis of variance (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 (Nunnally, 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. 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 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 *et al.*, 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 and Smith, 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 *et al.*, 2009). Hierarchical Bayes models can now be relatively easily implemented using existing statistical software, such as **BUGS** and **JAGS** (Lunn *et al.*, 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 alternatives, and/or the effort involved in programming alternative methods or interpreting their output. The present dissertation attempts to help remedy this undesirable state of affairs by developing 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.

One important application of hierarchical Bayes ANOVA is statistical mediation analysis in the social and behavioral sciences. The traditional statistical mediation analysis consists of three linear regression models proposed by Baron and Kenny (1986) which are introduced in Chapter 3 in detail. Thus, it inherits strong assumptions of the linear regression such as the Normal distribution of the continuous response and no heterogeneity among different populations. However, the same situation is that researchers in mediation analysis continue to use the traditional re-

gression models even though their data might violate the assumptions, for example, a discrete response variable or a hierarchical data structure. In this dissertation, we attempt to overcome these limitations by introducing single-level and multilevel Bayesian mediation models in which the hierarchical Bayes ANOVA is a special case. All these models we developed are based on Bayesian inference and simulated by MCMC (Markov Chain Monte Carlo) methods which will be first reviewed in following sections.

1.2 Brief Review of Bayesian Statistical Methods

1.2.1 Bayes' Theorem

During the past decades, Bayesian inference has drawn great attention in the social and behavioral sciences (Rossi *et al.*, 2006). The Bayesian method which was first introduced in scientific research has been widely accepted and applied to problems in behavioral sciences because of the continuing development of more powerful computational methods. Besides, more and more available experimental data provides researchers and practitioners ample opportunities for Bayesian modeling.

The basic inference behind all Bayesian methods is intuitive: Bayes' Theorem. That is, after obtaining current data, knowledge on unknown parameters are updated according to newly obtained data. Unlike the frequentists' view, the philosophy behind Bayesian inference is that, the incorporated prior knowledge of unknown parameters also plays a very important role in estimations. In the Bayesian framework, the prior knowledge of an unknown parameter θ is represented by a prior

probability distribution, denoted by $p(\theta)$. The prior information is incorporated to update the unknown parameter θ through Bayes theorem:

$$p(\theta|data) = \frac{p(\theta, data)}{p(data)} = \frac{p(\theta)p(data|\theta)}{p(data)}, \quad (1.1)$$

where $p(\theta)$ denotes the prior distribution of the unknown parameter, $p(data|\theta)$ denotes the likelihood of the data given the parameter. $p(data)$ is the prior or marginal probability of data and considered as a normalizing factor. By using Bayes' Theorem, we obtain the probability distribution of the unknown parameter θ given the data, $p(\theta|data)$, which is called the posterior distribution. Since $p(data)$ is just a normalizing factor, the above equation can also be written as follows,

$$p(\theta|data) \propto p(\theta)p(data|\theta) \quad (1.2)$$

or using the following notation,

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood} \quad (1.3)$$

Thus, the posterior distribution is proportional to the product of the likelihood and the prior distribution up to a constant.

1.2.2 Bayes Estimates and Markov Chain Monte Carlo (MCMC)

As we can see from Bayes' Theorem, all information about a parameter given the data is contained in the posterior distribution. Thus the estimation result of the parameter should not just be a point estimate but the whole distribution which contains all necessary estimates of the parameters of the distribution, for example, the mean and also the variance of a normal distribution. The most convenient and

informative way to represent this information is to plot the posterior distribution of the parameter, which provides a direct visual sense of the estimations. Based on the estimated posterior distribution, we can also derive some statistics of interest, such as the posterior mean and variance,

$$\hat{\theta} = E(\theta|data) = \int \theta p(\theta|data) d\theta, \quad (1.4)$$

and,

$$Var(\theta|data) = \int (\theta - \hat{\theta})^2 p(\theta|data) d\theta. \quad (1.5)$$

The other important information which can also be obtained from the posterior distribution is the credible interval (CI). The $(1 - \alpha)\%$ credible interval is defined as $[q_{\alpha/2}, q_{1-\alpha/2}]$, where $q_{\alpha/2}$ denotes the $\alpha/2$ quantile of the posterior distribution. A 95% credible interval is $[q_{0.025}, q_{0.975}]$, for instance.

However, if the integrals in Equations (1.4) and (1.5) don't have closed forms, it is not straightforward to obtain the posterior distribution. In this case, simulation methods such as the Markov Chain Monte Carlo (MCMC) have been developed to solve this problem. The basic idea of MCMC is to construct a Markov chain whose limit distribution is exactly the posterior distribution that we are interested in. Two popular MCMC methods are the Gibbs sampler and the Metropolis-Hastings sampler:

The Metropolis-Hastings (M-H) algorithm produces a Markov chain, $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots\}$, of the parameter of interest θ , whose limiting equilibrium distribution is the kernel of the posterior distribution of interest, $\pi(\theta)$. The algorithm is as follows:

1. Initial value: $\theta^{(0)}$

2. Proposed move: $\theta^* \sim q(\theta^*|\theta^{(i-1)})$

3. Acceptance scheme:

$$\theta^{(i)} = \begin{cases} \theta^*, & \text{with prob. } \alpha. \\ \theta^{(i-1)}, & \text{with prob. } 1 - \alpha. \end{cases} \quad (1.6)$$

where

$$\alpha = \min \left\{ 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})} \right\}. \quad (1.7)$$

Different proposal functions $q(\theta^*|\theta^{(i-1)})$ produce different Markov chains but the same limiting distribution. Two typical proposal functions are the independence function $q(\theta^*|\theta^{(i-1)}) = q(\theta^*)$ and the random walk function $q(\theta^*|\theta^{(i-1)}) = q(|\theta^* - \theta^{(i-1)}|)$.

The Gibbs sampler is a special case of the Metropolis-Hastings. In step 2 of the M-H algorithm, if we let the proposal function be $\pi(\theta^*|\theta^{(i-1)})$, then the acceptance ratio in the above equation would be

$$\alpha = \min \left\{ 1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{\pi(\theta^{(i-1)}|\theta^*)}{\pi(\theta^*|\theta^{(i-1)})} \right\} = \min \left\{ 1, \frac{\pi(\theta^*, \theta^{(i-1)})}{\pi(\theta^{(i-1)}, \theta^*)} \right\} = 1. \quad (1.8)$$

Thus, if the conditional distributions of all components are standard, we can easily apply the Gibbs sampler instead of M-H sampling. The general steps of the Gibbs sampler are: given $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_p^{(t)})$, generate

1. $\theta_1^{(t+1)} \sim \pi_1(\theta_1|\theta_2^{(t)}, \dots, \theta_p^{(t)})$
2. $\theta_2^{(t+1)} \sim \pi_2(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)})$
- ...
- p. $\theta_p^{(t+1)} \sim \pi_p(\theta_p|\theta_1^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})$

where $\theta_i^{(t)}$, $i = 1, \dots, p$, are the parameters of interest at each iteration t . The π_i , $i = 1, \dots, p$, are conditional posterior distributions of parameters θ_i , $i = 1, \dots, p$, given remaining parameters. From the above steps, we can see that, at each iteration, the Gibbs sampler samples one parameter conditioning on other updated parameters sequentially until all distribution converge.

There are a lot of comprehensive reviews of MCMC simulation method (such as Robert and Casella (2004)). Thus we do not include every detail here. The MCMC sampling method is now widely used and considered as the solution for Bayesian computation.

1.3 Overview

In our work presented in the following chapters, we first (in Chapter 2) introduce a new R package called **BANOVA** developed by Dong and Wedel (2014b) which combines hierarchical Bayesian linear regression and analysis of variance (ANOVA) techniques. Several data sets (Etkin and Ratner (2012), Ferraro *et al.* (2013), Wedel and Pieters (2014)) are used as examples to illustrate the applications of the models we developed.

The hierarchical Bayesian ANOVA models in the R package deal with a wide range of distributions for the dependent variable (including Normal, Student's t , Poisson, Bernoulli, Binomial, ordered and unordered Multinomial distributions), with hierarchical data structures and between- and within- subjects design factors, as well as continuous covariates. The combination of generalized linear models in

subject-level (within subjects), ANOVA models in population-level (between subjects), and Bayesian sampling methods can help to analyze any kind of multilevel data and obtain the effects of all parameters in a very straightforward way.

In Chapter 3, we introduce new Bayesian models for mediation analysis (Dong and Wedel, 2014a). By using Bayesian sampling methods, the models don't require sophisticated and strict assumptions to test the indirect effect, since its posterior distribution can be directly obtained. We illustrate the applications of our Bayesian mediation models using both single-level and multilevel data sets, for which similar or even better statistical results are obtained.

Finally, we discuss potential extensions of our models and future research work to be done in Chapter 4.

Chapter 2: **BANOVA**: An R Package for Hierarchical Bayesian ANOVA

2.1 Introduction

In this chapter, we introduce an R package for hierarchical Bayes ANOVA which combines both hierarchical Bayes and ANOVA naturally. 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 chapter 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.

2.2 Bayesian ANOVA Models

We assume data are collected in an experiment in which a sample of subjects have participated, 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, μ_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 ($p = 1, 2, \dots, P$). Each index p represents a batch of J_p coefficients: $\beta_{j,s}^p$, $j = 1, \dots, J_p$; $s = 1, \dots, S$ indexes subjects. Note that if a subject-level covariate is continuous, $J_p = 1$, so that ANCOVA models are also accommodated (but the formulation here relaxes their “constant slope” assumption). The subject-level model is expressed as a generalized linear regression model, with a design matrix X that contains all within-subject factors and their interactions, as well as a constant term ($p = 0$):

$$\mathbf{E}(y_i) = g^{-1}(\eta_i), \tag{2.1}$$

$$\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p, \tag{2.2}$$

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 $\beta_{j,s}^p$ are assumed to follow a multivariate normal distribution. The between-subject factors and their interactions are indexed by q ($q = 1, 2, \dots, Q$); $q = 0$ denotes the constant term. Then, using the notation in

Gelman (2005), the population-level ANOVA can be written as:

$$\beta_{j,s}^p = \sum_{q=0}^Q \theta_{j,k_s^q}^{p,q} + \delta_{j,s}^p. \quad (2.3)$$

Each q represents a batch of K_q coefficients: $\theta_{j,k}^{p,q}$, $k = 1, \dots, K_q$; K_s^q 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 ($P = 2$, $J_1 = 1$ and $J_2 = 3$) and two 2-level between-subject factors A and B , and the AB -interaction, $Q = 3$, equation (2.2) and equation (2.3) reduce to (with the parameter of the last level of each factor set to zero):

$$\eta_i = \beta_{1,s}^0 + \sum_{j=1}^2 X_{i,j}^D \beta_{j,s_i}^D, \quad (2.4a)$$

$$\beta_{1,s}^0 = \theta_1^0 + \theta_{1,k_s^A}^A + \theta_{1,k_s^B}^B + \theta_{1,k_s^{AB}}^{AB}, \quad (2.4b)$$

$$\beta_{1,s}^D = \theta_1^{D0} + \theta_{1,k_s^A}^{DA} + \theta_{1,k_s^B}^{DB} + \theta_{1,k_s^{AB}}^{DAB}, \quad (2.4c)$$

$$\beta_{2,s}^D = \theta_2^{D0} + \theta_{2,k_s^A}^{DA} + \theta_{2,k_s^B}^{DB} + \theta_{2,k_s^{AB}}^{DAB}. \quad (2.4d)$$

Here, equation (2.4b) contains the overall intercept (θ_1^0) and the main effects of the between-subject factors A ($\theta_{1,k_s^A}^A$) and B ($\theta_{1,k_s^B}^B$) and their interaction ($\theta_{1,k_s^{AB}}^{AB}$). Equation (2.4c) contains the main effect of the (first level) of the within-subject factor D (θ_1^{D0}), and its two- and three-way interactions with the between-subject factors and B. Similarly, equation (2.4d) 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):

$$\beta_{j,s}^p = \sum_{k=1}^K Z_{s,k} \theta_{j,k}^p + \delta_{j,s}^p, \quad (2.5)$$

where $Z_{s,k}$ is an element of Z , a $S \times K$ matrix of covariates, K is the number of parameters. $\theta_{j,k}^p$ is a hyper-parameter which captures the effects of between-subject factor on the parameter $\beta_{j,s}^p$ of within-subject factor p . The error $\delta_{j,s}^p$ is assumed to be normal: $\delta_{j,s}^p \sim N(0, \sigma_p^{-2})$. Proper, but diffuse priors are assumed: $\theta_{j,k}^p \sim N(0, s)$, and $\sigma_p \sim \text{Gamma}(a, b)$, 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 (2.1), (2.2) and (2.5). We use effects coding of the factors in the matrices X and Z (Overall *et al.*, 1975). It is important to note that equation (2.3) 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 $\beta_{j,s}^p$ are the (Normally distributed) “dependent variables” in an ANOVA, specified by the between-subject factors in equation (2.3). Thus, inference focuses on the parameters in equation (2.3). 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 :

$$y_i = \eta_i + \epsilon_i; \quad \epsilon_i \sim N(0, \sigma^{-2}), \quad (2.6)$$

where η_i is defined in equation (2.2), and the scale parameter σ , $\sigma \sim \text{Gamma}(\alpha, \beta)$.

To describe data with “outliers” or fatter tails than the normal, the distribution

of ϵ_i is assumed to follow a t -distribution, with an unknown number of degrees of freedom, assumed to follow a Poisson distribution:

$$\epsilon_i \sim t(0, \phi^{-2}, \nu), \phi \sim \text{Gamma}(\alpha, \beta), \nu \sim \text{Poisson}(\lambda), \quad (2.7)$$

where α, β, λ are hyper-parameters.

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

$$y_i \sim \text{Binomial}(1, p_i), p_i = \text{logit}^{-1}(\eta_i), \quad (2.8)$$

where $\text{logit}(x) = \log[x/(1-x)]$ is the standard logit link-function.

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

$$y_i \sim \text{Binomial}(B, p_i), p_i = \text{logit}^{-1}(\eta_i), \quad (2.9)$$

Count responses: To model count data y_i that can take on values in $0, 1, 2, \dots$, the Poisson distribution is assumed:

$$y_i \sim \text{Poisson}(\lambda_i); \lambda_i = \exp(\eta_i). \quad (2.10)$$

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

$$\Pr(y_i > m) = \text{logit}^{-1}(\eta_i - c_{m-1}); m = 1, \dots, M - 1. \quad (2.11)$$

The cut-point parameters c_k are constrained: $0 = c_1 < c_2 < \dots < c_{K-1}$. We assume $c_1 = 0$, and the other cut-points are the order statistics of $\hat{c}_2, \hat{c}_3, \dots, \hat{c}_{K-1}$ where each follows a uniform distribution $c_m \sim \text{Uniform}(0, 10)$.

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

$$Pr(y_i = k) = \frac{\exp(\eta_{i,k})}{\sum_{k=1}^K \exp(\eta_{i,k})} \quad (2.12)$$

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

2.3 BANOVA R Package

2.3.1 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/>). In addition, a web application based on this package is also developed and introduced in Appendix C. 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 *et al.*, 2006) in order to connect R and **JAGS** and to perform convergence diagnostics. Note that, 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> library('BANOVA')
```

2.3.2 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 2.8).
2. `BANOVA.Bin()`—the model in which the response variable follows a binomial distribution (equation 2.9).
3. `BANOVA.Multinomial()`—the model in which the response variable follows an unordered multinomial distribution (equation 2.12).
4. `BANOVA.Normal()`—the model in which the response variable follows a normal distribution (equation 2.6).
5. `BANOVA.ordMultinomial()`—the model in which the response variable fol-

lows an ordered categorical distribution (equation 2.11).

6. `BANOVA.Poisson()`—the model in which the response variable is considered a count variable which follows a Poisson distribution (equation 2.10).

7. `BANOVA.T()`—the model in which the response variable follows a t distribution (equation 2.7).

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. A complete manual of the package is included in Appendix B.

2.3.3 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 trail, 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 2.1), 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> class(x) # will display 'integer' number of classes in variable x
R> x <- as.factor(x) # class of x is changed to 'factor'
```

Subject_ID	Category	Brand_Accuracy	Size	Blur	Color
1	1	1	0	0	0
1	1	0	0	0	0
1	1	1	0	0	0
1	2	0	0	0	0
1	2	1	0	0	0
1	2	1	0	0	0
2	2	0	0	1	1
2	2	1	0	1	1
2	2	1	0	1	1
2	1	0	0	1	1
2	1	1	0	1	1
3	1	0	1	0	1
3	1	0	1	0	1
3	1	1	1	0	1
3	2	0	1	0	1
3	2	0	1	0	1
3	2	1	1	0	1
4	2	0	1	1	0
4	2	1	1	1	0
4	2	0	1	1	0
4	1	1	1	1	0
4	1	0	1	1	0
4	1	1	1	1	0

Figure 2.1: Example of the format of input data

For the multinomial response model (equation 2.12), 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> 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]
```

2.3.4 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 (2.1), (2.2) and (2.5). The conditional posterior distribution, denoted by $\pi(\cdot)$, of the parameters $\beta_{j,s}^p$ is obtained from the likelihood and priors:

$$\pi(\beta_{j,s}^p|\cdot) \propto \pi(y_i|\beta^1, \dots, \beta^p, x_i)\pi(\beta^1|\theta^1, \mathbf{Z}) \cdots \pi(\beta^p|\theta^p, \mathbf{Z})\pi(\theta^1) \cdots \pi(\theta^p), \quad (2.13)$$

where $\pi(y_i|\beta^1, \dots, \beta^p, x_i)$ is the likelihood determined by equations (2.1), (2.2) and (2.6) to (2.10); $\beta^p = (\beta_1^p, \dots, \beta_{J_p}^p)$, $p = 1, \dots, P$, with $\beta_j^p = (\beta_{j,1}^p, \dots, \beta_{j,s}^p)$, $j = 1, \dots, J_p$,

are the coefficients of factor p ; $\boldsymbol{\theta}^p = (\theta_{j,k}^{p1}, \theta_{j,k}^{p2}, \dots, \theta_{j,k}^{pQ})$, $j = 1, \dots, J_p, k = 1, \dots, K_q$ are the population-level parameters; $\pi(\boldsymbol{\theta}^p)$, $p = 1, \dots, P$ is the prior of population-level parameters $\boldsymbol{\theta}^p$; and $\pi(\boldsymbol{\beta}^p | \boldsymbol{\theta}^p, \mathbf{Z})$, $p = 1, \dots, P$, is the prior determined by the population model in equation (2.3). We are interested in the effects of between-subject level factors, captured by $\theta_{j,k}^{pq}$, $k = 1, \dots, K_q$ using the notation in equation (2.3). The conditional posterior distribution of the parameters $\theta_{j,k}^{pq}$ is:

$$\begin{aligned} \pi(\theta_{j,k}^{pq} | \cdot) &\propto \pi(\boldsymbol{\beta}_j^p | \boldsymbol{\theta}^p, \mathbf{Z}) \pi(\boldsymbol{\theta}^p) \\ &= \pi(\boldsymbol{\beta}_j^p | \boldsymbol{\theta}^p, \mathbf{Z}) \prod_{jk} \pi(\theta_{j,k}^{p1}) \prod_{jk} \pi(\theta_{j,k}^{p2}) \cdots \prod_{jk} \pi(\theta_{j,k}^{pQ}). \end{aligned} \quad (2.14)$$

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

2.3.5 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 $\theta_{j,k}^{pq}$ in equation (2.14), for example). For those parameters with constraints, for instance, the hyper-parameters α, β, ν in equation (2.7), are assigned fixed values ($\alpha = 1, \beta = 1, \nu = 1$). These values can also be found through the corresponding **JAGS** code.

2.3.6 JAGS Code

Based on equation (2.13) and (2.14), 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 an 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:

```
R> cat(res$JAGSmodel) #res is a list returned from the BANOVA.* function
```

2.3.7 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 the 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 $\frac{m_A}{m}$ and $\frac{m_B}{m}$ 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 process. 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:

```
R> conv.diag(res) // res is a fitted object from any of the models
```

2.3.8 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 $\theta_{j,k}^{pq}$ in equation (2.3) 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 $\theta_{j,k,m}^{pq}$ denote the posterior sample of $\theta_{j,k}^{pq}$ in the m th iteration of the MCMC chain. Then the grand mean is:

$$\bar{\mu} = \frac{\sum_m g^{-1}(\theta_m^{00})}{M}, \quad (2.15)$$

where θ_m^{00} 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 $\{g^{-1}(\theta_i^{00})\}$, 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:

$$\bar{\mu}_j^A = \frac{\sum_m g^{-1}(\theta_m^{00} + \mathbf{X}_j^{A'} \boldsymbol{\theta}_m^{A0})}{M}, \quad (2.16)$$

where \mathbf{X}_j^A 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; $\boldsymbol{\theta}_m^{A0}$ 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:

$$\bar{\mu}_j^B = \frac{\sum_m g^{-1}(\theta_m^{00} + \mathbf{Z}'_{j,B} \boldsymbol{\theta}_m^{0B})}{M}, \quad (2.17)$$

where $\mathbf{Z}_{j,B}$ 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; $\boldsymbol{\theta}_m^{0B}$ 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:

$$\bar{\mu}_{j,k}^{AB} = \frac{\sum_m g^{-1}(\theta_m^{00} + \mathbf{X}_j^{A'} \boldsymbol{\theta}_{j,k,m}^{AB} \mathbf{Z}'_{k,B})}{M}, \quad (2.18)$$

where $\boldsymbol{\theta}_{j,k,m}^{AB}$ 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
```

2.3.9 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 (2.3) or (2.5) 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 (2.5), so that for $\boldsymbol{\theta}_j^p = (\theta_{j,1}^p, \theta_{j,2}^p, \dots, \theta_{j,K}^p)$, the total sum of squares can be represented as

$$SS(\boldsymbol{\theta}_j^p) = \boldsymbol{\theta}_j^{p'} \mathbf{Z}' \boldsymbol{\beta}_j^p. \quad (2.19)$$

In the package, equation (2.19) is estimated as $SS(\boldsymbol{\theta}_j^p) = (1/M) \sum \boldsymbol{\theta}_{j,m}^{p'} \mathbf{Z}' \boldsymbol{\beta}_{j,m}^p$ where m indexes the posterior samples of $\boldsymbol{\theta}_j^{p'}$ and $\mathbf{Z}' \boldsymbol{\beta}_j^p$. If the design matrix \mathbf{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 \mathbf{Z}_q of \mathbf{Z} corresponding to each factor q and its coefficients $\theta_j^{pq} = (\theta_{j,1}^{pq}, \dots, \theta_{j,K_q}^{pq})$, $q = 1, \dots, Q$. (e.g. Draper and Smith, 1998),

$$SS(\theta_j^{pq}) = \boldsymbol{\theta}_j^{pq'} \mathbf{Z}'_q \boldsymbol{\beta}_j^p, \quad (2.20)$$

where $\boldsymbol{\theta}_j^{pq}$ is the vector of coefficients of the dummy variables corresponding to factor or interaction q . Equation (2.20) is estimated as $SS(\theta_j^{pq}) = (1/M) \sum \boldsymbol{\theta}_{j,m}^{pq'} \mathbf{Z}'_q \boldsymbol{\beta}_{j,m}^p$ in the package. For orthogonal designs it holds that

$$SS(\boldsymbol{\theta}_j^p) = SS(\theta_j^{p1}) + SS(\theta_j^{p2}) + \dots + SS(\theta_j^{pQ}). \quad (2.21)$$

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 (η_p^2), is calculated. It is defined as

$$\eta^2 = \frac{SS_{effect}}{SS_{total}}, \quad (2.22)$$

where SS_{effect} = the sum of squares for the effect of interest, SS_{total} = the total sum of squares for all effects, interactions, and errors in the regression.

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

```
R> BAnova(res) // res is a fitted object from any of the models
```

2.3.10 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 $\theta_{j,k}^{pq}$ in equation (2.3) are

$$H_0 : \theta_{j,k}^{pq} = 0, \text{ versus } H_1 : \theta_{j,k}^{pq} \neq 0. \quad (2.23)$$

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

$$\min\left(\frac{\sum_m (I(\theta_{j,k,m}^{pq} < 0))}{M}, \frac{\sum_m (I(\theta_{j,k,m}^{pq} > 0))}{M}\right). \quad (2.24)$$

Then, the two sided p-value is

$$P_\theta(\theta_{j,k}^{pq}) = 2 \times \min(P(\theta_{j,k}^{pq} < 0), P(\theta_{j,k}^{pq} > 0)). \quad (2.25)$$

If there are coefficients $\theta_{j,k_1}^{pq}, \theta_{j,k_2}^{pq}, \dots, \theta_{j,k_J}^{pq}$ 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:

$$H_0 : \theta_{j,k_1}^{pq} = \theta_{j,k_2}^{pq} = \dots = \theta_{j,k_J}^{pq} = 0, \text{ versus } H_1 : \text{some } \theta_{j,k_j}^{pq} \neq 0. \quad (2.26)$$

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

$$\min(P_\theta(\theta_{j,k_{min}}^{pq}), P_\theta(\theta_{j,k_{max}}^{pq})). \quad (2.27)$$

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
```

2.4 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 *et al.* (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.

2.4.1 Application 1: Impact of the Variety among Means on Motivation

In this examples we illustrate the application of the **BANOVA** package to 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> data(goalstudy)
```

The structure of the data is shown below:

```
R> head(goalstudy)
```

	id	perceivedsim	goalprogress	varmeans	bid
1	1	5	1	2	5
2	2	7	1	1	0
3	3	2	2	2	1
4	4	2	2	1	15
5	5	5	2	1	3
6	6	5	1	1	10

Table 2.1: Sample of the `goalstudy` data

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 ($\log(\text{bid} + 1)$) 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> goalstudy$logbid <- log(goalstudy$bid + 1)
R> app_1 <- BANOVA.Normal(logbid~1, ~goalprogress*varemeans, 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) : varemeans1	Passed	0.6585
(Intercept) : goalprogress1:varemeans1	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 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 below.

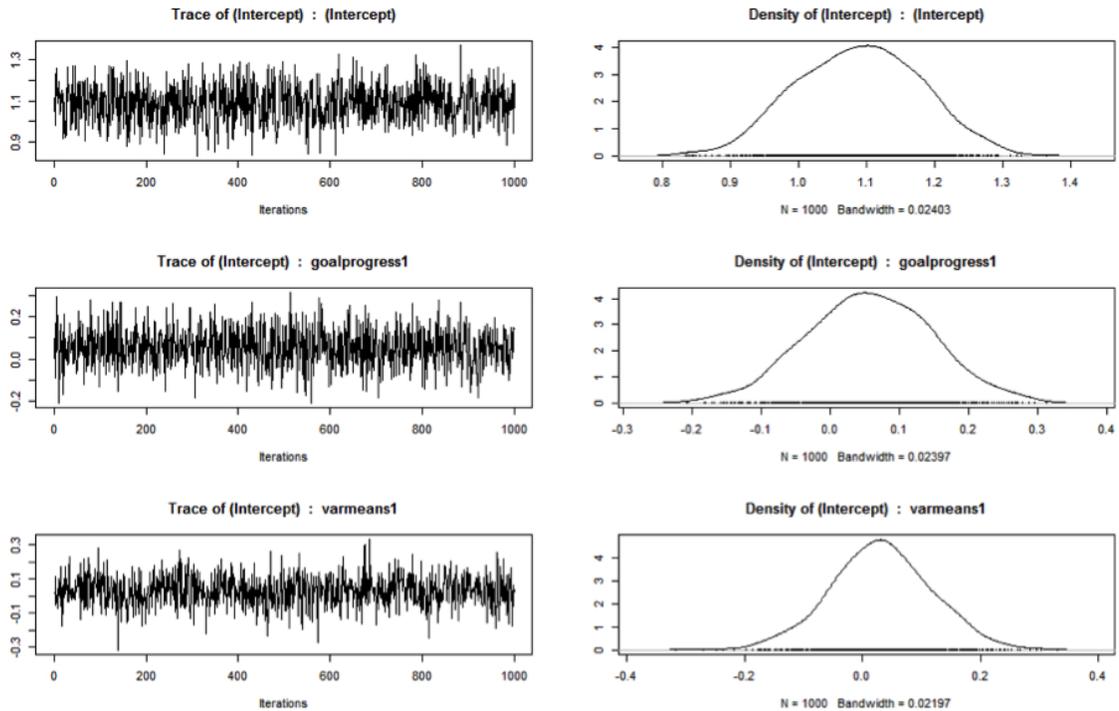


Figure 2.2: Trace plots of (selected) posterior distributions of the parameters of the goalstudy application

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> 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):

	goalprogress	varmeans	goalprogress.varmeans	Residuals	Total
(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)	: (Intercept)	1.08841	0.09027	0.91787	1.26383	<2e-16	***
(Intercept)	: goalprogress1	0.056	0.09002	-0.12835	0.23335	0.538	
(Intercept)	: varmeans1	0.03147	0.08717	-0.14056	0.20189	0.706	
(Intercept)	: goalprogress1:varmeans1	-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 results above, the interaction between variety among means (varmeans) and goal progress (goalprogress) is significant. The table of means, produced with the command shown 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> table.means(app_1)
```

Grand mean:

1.08841

2.5% 97.5%

0.91787 1.26383

Means for factors at level 2:

goalprogress mean 2.5% 97.5%

1 1.15012 0.88968 1.39588

2 1.02906 0.79407 1.27638

varmeans mean 2.5% 97.5%

1 1.12212 0.87387 1.35245

2 1.05696 0.7941 1.30155

Means for interactions at level 2:

goalprogress varmeans mean 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> # 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> # 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> 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):

	(Intercept)	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	0.25257	0.87931	0.002	**
(Intercept)	: goalprogress1	0.08994	0.14916	-0.18453	0.39551	0.53	
(Intercept)	: varmeans1	0.06434	0.14453	-0.21212	0.34755	0.672	
(Intercept)	: goalprogress1:varmeans1	-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, which is expected to depend upon 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, and the results are provided below.

```
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):
```

	(Intercept)	goalprogress	varmeans	goalprogress:varmeans
(Intercept)	0	0.51	0	0.332

```
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:varmeans1	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>
Cutpoint[6]	8.741805	0.953105	6.546533	9.959254	<NA>	<NA>

```
---
```

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

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> table.means(app_2)
```

```
Table of means of the response
```

```
-----
```

```
Grand mean:
```

```
4.23097
```

```
2.5% 97.5%
```

```
3.87904 4.57188
```

```
Means for factors at level 2:
```

```
goalprogress mean 2.5% 97.5%
```

```
1 4.1277 3.60441 4.5929
```

```
2 4.32479 3.85958 4.7886
```

```
varmeans mean 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 mean 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:

0.00915

2.5% 97.5%

0.00284 0.03222

Means for factors at level 2:

goalprogress mean 2.5% 97.5%

1 0.01125 0.00297 0.04087

2 0.00807 0.00218 0.03081

varmeans mean 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	mean	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	1	0.05626	0.01281	0.18767
[2,]	1	2	0.27994	0.10976	0.50518
[3,]	1	3	0.30298	0.16622	0.46136
[4,]	1	4	0.20120	0.08078	0.37816
[5,]	1	5	0.09312	0.02778	0.24850
[6,]	1	6	0.01735	0.00352	0.07083
[7,]	1	7	0.00267	0.00045	0.02066
[8,]	2	1	0.00206	0.00043	0.01556
[9,]	2	2	0.01729	0.00368	0.07122
[10,]	2	3	0.05283	0.01322	0.14533
[11,]	2	4	0.15319	0.05419	0.30549
[12,]	2	5	0.39464	0.24039	0.54825
[13,]	2	6	0.27310	0.10555	0.47190
[14,]	2	7	0.07001	0.01975	0.20811

The **JAGS** code for the above model generated by the program can be easily retrieved by the R command shown in Figure 2.3.

```

R> cat(app_2$JAGSmodel)

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)
}

```

Figure 2.3: **JAGS** code for the `goalstudy` application with `perceivedsim` as the dependent variable

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.

2.4.2 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 second 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 treated as a continuous dependent variable, assumed to follow a Normal distribution. Data from 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_d1
1	1	3.000000	0	19	0	0	-2.3042672	3
2	2	5.333333	0	33	0	1	1.6957328	6
3	3	5.666667	0	25	1	0	-0.1614100	6
4	4	5.333333	1	41	0	1	-0.4471243	5
5	5	6.000000	1	38	1	1	0.2671614	6
6	6	4.000000	0	33	1	0	0.6957328	4

Table 2.2: Sample of the `ipadstudy` data

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	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	-0.0204	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 <- BANOVA.T(attitude~1, ~owner + age + gender + selfbrand*conspic,
+                   ipadstudy, ipadstudy$id, burnin = 3000, sample = 1000, thin = 5 )
R> summary(app_4)
```

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 (3.03%)	0.34 (0.09%)	0.354 (0.1%)	126.2 (34.91%)	9.01 (2.49%)
	selfbrand.conspic	Residuals	Total		
	6.251 (1.73%)	135.715	361.517		

Table of p-values (Multidimensional):

	(Intercept)	owner	age	gender	selfbrand	conspic	selfbrand:conspic
(Intercept)		0	0	0.942	0.94	0	0.01
							0.028

Table of coefficients:

	mean	SD	Quantile0.025	Quantile0.975	p.value	Signif. codes
(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

2.4.3 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 Gaussian blur filters of varying 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
```

Table 2.3: Sample of the `colorad` data

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`, `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.

```
R> summary(app_5)
```

Call:

```
BANOVA.Bin(l1_formula = y ~ typic, l2_formula = ~color * blur,
  data = colorad, id = colorad$id, num_trails = as.integer(16),
  burnin = 3000, sample = 2000, thin = 5)
```

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

	color	blur	color.blur	Residuals	Total
(Intercept)	4.691 (2.85%)	26.344 (15.99%)	1.679 (1.02%)	131.733	164.733
typic1	1.55 (10.36%)	5.651 (37.76%)	1.403 (9.38%)	6.307	14.965

Table of p-values (Multidimensional):

	(Intercept)	color	blur	color:blur
(Intercept)	0	0.011	0	0.166
typic	0	0.003	0	0.011

Table of coefficients:

		mean	SD	Quantile0.025	Quantile0.975	p.value	Signif.codes
(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	***
(Intercept)	: color1:blur	0.04027	0.02917	-0.01769	0.09815	0.166	
typic1	: (Intercept)	0.30226	0.02707	0.25178	0.35805	<2e-16	***
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	: color1:blur	0.03697	0.01441	0.00895	0.06458	0.011	*

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:

```
R> app_6 <- BANOVA.Bin(y~typic, ~color*blurfac, colorad, colorad$id,
+ 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 `summary()`, the results are presented below.

```
R> summary(app_6)
```

Call:

```
BANOVA.Bin(l1_formula = y ~ typic, l2_formula = ~color * blurfac,
  data = colorad, id = colorad$id, num_trails = as.integer(16),
  burnin = 20000, sample = 3000, thin = 5)
```

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

	color	blurfac	color.blurfac	Residuals	Total
(Intercept)	5.796 (3.43%)	31.149 (18.42%)	3.284 (1.94%)	128.956	169.139
typic1	2.808 (14.35%)	14.029 (71.68%)	1.753 (8.96%)	1.112	19.571

Table of p-values (Multidimensional):

	(Intercept)	color	blurfac	color:blurfac
(Intercept)	0	0.0087	0	0.1547
typic	0	0.0000	0	0.0140

Table of coefficients:

	mean	SD	Quantile0.025	Quantile0.975	p.value	Signif.codes
(Intercept) : (Intercept)	0.49252	0.05571	0.38637	0.60383	<2e-16	***
(Intercept) : color1	0.15379	0.05712	0.04019	0.26625	0.0087	**
(Intercept) : blurfac1	0.55405	0.11197	0.33141	0.77124	<2e-16	***

(Intercept)	:	blurfac2	0.0506	0.11644	-0.18528	0.28439	0.6507	
(Intercept)	:	blurfac3	0.01659	0.10501	-0.18748	0.22031	0.882	
(Intercept)	:	blurfac4	-0.05149	0.10783	-0.25875	0.16721	0.6313	
(Intercept)	:	color1:blurfac1	-0.15806	0.10887	-0.3719	0.05631	0.1547	
(Intercept)	:	color1:blurfac2	0.12066	0.11757	-0.10787	0.34741	0.3073	
(Intercept)	:	color1:blurfac3	-0.08139	0.10941	-0.29386	0.12871	0.4533	
(Intercept)	:	color1:blurfac4	0.00974	0.10998	-0.21415	0.21614	0.9173	
typic1	:	(Intercept)	0.32106	0.02758	0.25963	0.35879	<2e-16	***
typic1	:	color1	0.1043	0.0323	0.04425	0.14222	<2e-16	***
typic1	:	blurfac1	0.17167	0.04963	0.06052	0.26182	0.002	**
typic1	:	blurfac2	0.36932	0.04478	0.27415	0.46363	<2e-16	***
typic1	:	blurfac3	-0.05029	0.05149	-0.12272	0.05379	0.3847	
typic1	:	blurfac4	-0.17165	0.04197	-0.24914	-0.07217	0.002	**
typic1	:	color1:blurfac1	-0.11062	0.0554	-0.22147	-0.03255	0.014	*
typic1	:	color1:blurfac2	0.04417	0.05024	-0.06631	0.13594	0.3867	
typic1	:	color1:blurfac3	-0.03529	0.05048	-0.10884	0.07542	0.5107	
typic1	:	color1:blurfac4	0.09386	0.04141	0.00489	0.17845	0.032	*

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

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> table.means(app_6)
```

```
Grand mean:
```

```
0.6206999
```

```
      2.5%    97.5%
```

```
0.5954085 0.6465321
```

```
Means for factors at level 1:
```

```
typic mean    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 mean    2.5%    97.5%
```

```
0    0.65594 0.61981 0.69103
```

```
1    0.58403 0.54472 0.62252
```

```
blurfac mean  2.5%    97.5%
```

```
1    0.74044 0.68921 0.78352
```

```
2    0.63253 0.56918 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	mean	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	mean	2.5%	97.5%
0	1	0.82398	0.77602	0.86012
0	2	0.77421	0.71804	0.8217
0	3	0.68515	0.62832	0.737
0	4	0.64344	0.58566	0.69783
0	5	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	mean	2.5%	97.5%
0	1	0.73962	0.66789	0.79778
0	2	0.69298	0.60684	0.76707
0	3	0.6415	0.56029	0.71427
0	4	0.64691	0.56745	0.71823
0	5	0.54569	0.46092	0.6329
1	1	0.74048	0.66902	0.80171
1	2	0.56669	0.47772	0.65084
1	3	0.60753	0.52846	0.6833
1	4	0.56849	0.48956	0.6494

1 5 0.41635 0.33232 0.50044

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

typic	color	blurfac	mean	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	0.38974	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 (Ttypic= 1, color = 1, blur = 1,...,5) when there is no blur, or a low level of blur (Wedel and Pieters, 2014).

2.5 Conclusion

This chapter 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 models, including **BACC** (Geweke, 1999), **bayesm** (Rossi and McCulloch, 2006) and **MCMCpack** (Martin *et al.*, 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.

Chapter 3: Bayesian Mediation Analysis

3.1 Introduction

Mediation analysis is a statistical method used to investigate the relationship between an independent variable and a dependent variable via the inclusion of a third explanatory variable, known as a mediator variable. In a mediational structure, an independent variable might affect a dependent variable not directly but rather through a third variable which plays an important role in governing the relationship between the other two variables. For example, attitude toward an advertisement may enhance attitude toward a brand, which in turn may positively impact likelihood to purchase the brand.

Mediation has been of interest to many social science researchers, including Baron and Kenny (1986), Iacobucci (2001), James and Brett (1984) and MacKinnon *et al.* (2002). The researchers focus their studies on better understanding whether the effect of the independent variable on the dependent variable is direct or indirect via the mediator. A simple statistical mediation model is described in Figure 3.1.

As depicted in Figure 3.1, X is the independent variable; M , the hypothesized mediator; and Y , the dependent variable. A simple mediation model is any causal system in which at least one causal independent variable X is proposed as influenc-

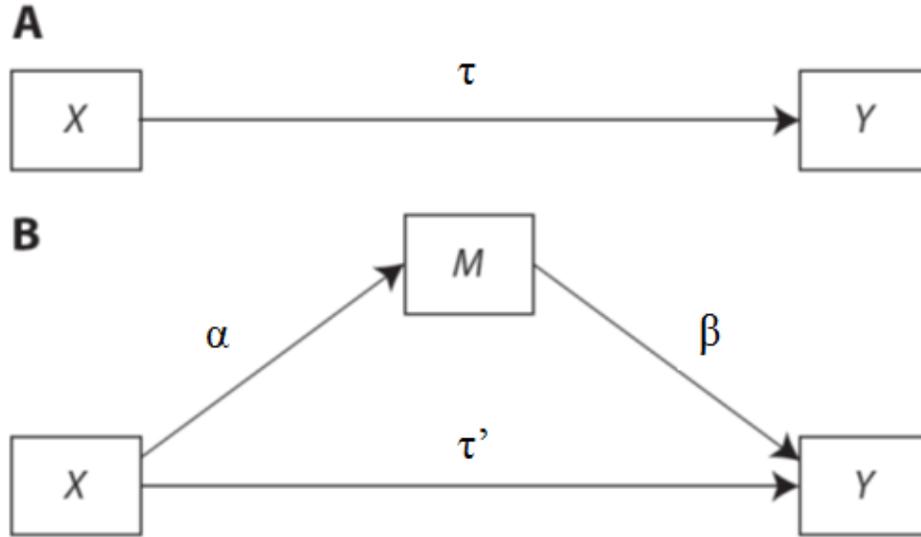


Figure 3.1: A diagram of a simple mediation model.

ing an outcome Y through a single or multiple mediators M . In studies in which mediation is posited and tested, the question is whether the direct effect of X on Y , $X \rightarrow Y$ is significant or mediated through the mediator variable M , $X \rightarrow M \rightarrow Y$.

Regression based approaches are usually used by researchers to test these relationships empirically. The most popular means of testing for mediation is the procedure proposed by Baron and Kenny (1986) in which a series of three regular regression models are fitted:

$$Y = \beta_1 + \tau X + \epsilon_1, \quad (3.1)$$

$$M = \beta_2 + \alpha X + \epsilon_2, \quad (3.2)$$

$$Y = \beta_3 + \tau' X + \beta M + \epsilon_3, \quad (3.3)$$

where the intercepts $\beta_1, \beta_2, \beta_3$ and the regression errors $\epsilon_1, \epsilon_2, \epsilon_3$ are usually ignored.

The regression coefficients α, β, τ and τ' capturing the relationships between the

three focal variables are usually of interest. Evidence for mediation is said to be likely if

1. The independent variable is a significant predictor of the dependent variable.

That is, the coefficient τ is significant in equation (3.1).

2. The independent variable is a significant predictor of the mediator. That is, the coefficient α is significant in equation (3.2).

3. The mediator is a significant predictor of the dependent variable, while controlling for the independent variable. That is, the coefficient β is significant and τ' should be smaller in absolute value than the original effect τ in equation (3.3).

Based on these steps, considerable research has been conducted in mediation analysis: MacKinnon and Dwyer (1993); MacKinnon *et al.* (2000, 2002); Kraemer *et al.* (2002); Shrout and Bolger (2002); Zhao *et al.* (2010), as well as others. These works mainly focus on the single-level mediation model in which variables are mainly continuous. For mediation analysis with categorical variables, researchers have proposed various solutions to working with categorical variables or a mix of categorical and continuous variables. For example, Hayes and Preacher (2011) focus on X and allow it to be multinomial, not just binary. Iacobucci (2012) has proposed a single solution to all possible combinations of X , M , and Y being continuous or categorical. In the recent literature, there is growing interest in multilevel mediation analysis that is useful for analyzing data sets with hierarchical structure or repeated measures, for example, the research of Kenny *et al.* (1998), Krull and MacKinnon (1999, 2001), Raudenbush and Sampson (1999), Kenny *et al.* (2003), Bauer *et al.* (2006). However, all these researchers focus on mediation analysis from the frequen-

tist perspective. To date, very little research has been based on Bayesian techniques. For example, Yuan and MacKinnon (2009) proposed Bayesian analysis of mediation effects in both single-level and multilevel models, but their work was limited to continuous variables and subject-level models. Zhang *et al.* (2009) present the first mediation analysis in marketing properly estimated using Bayesian techniques with latent instrumental variables, to account for endogeneity.

In this chapter, we propose a Bayesian analysis of mediation effects which can be considered as a single Bayesian solution to all possible combinations of X , M , and Y being continuous or categorical in single-level or multilevel models. Compared with conventional frequentist mediation analysis, the Bayesian approach has several advantages. First, the MCMC estimation of Bayesian mediation models allows for simulation and computation of the posterior distribution and the standard error of a mediated effect. Based on posterior samples, credible intervals can be easily constructed and hypothesis testing can also be conducted in a straightforward manner. Second, researchers can choose appropriate priors for the mediation analysis, which may greatly improve the accuracy of estimations. Third, hierarchical Bayesian models introduced by (Gelman and Hill, 2007) and software, such as **WinBUGS**, can be used to build and solve mediation multilevel models in an easy and natural way. Using Bayesian methods, parameters in almost any multilevel model can be estimated without much difficulty. On the other hand, classical methods used to solve multilevel models often present difficulties in estimation (Kenny *et al.*, 2003).

3.2 Single Level Mediation Models

3.2.1 Conventional Mediation Analysis of Single-Level Models

The simple mediation model with one continuous independent variable X , mediator M and dependent variable Y was introduced in equations (3.1) to (3.3) and Figure 3.1. It is assumed that ϵ_1, ϵ_2 and ϵ_3 follow normal distributions with mean 0 and variances σ_1^2, σ_2^2 and σ_3^2 , respectively. For a mediated effect, it is clear that

$$\tau - \tau' = \alpha\beta. \quad (3.4)$$

Thus, there are two ways to calculate the mediated effect. The first one uses the term on the left, which is computed as the difference of two maximum-likelihood estimates, $\hat{\tau} - \hat{\tau}'$. The second one calculates the product of two estimated regression coefficients $\hat{\alpha}\hat{\beta}$ of α and β in equations (3.2) and (3.3). However, the Equation 3.4 only holds in the case of simple linear regressions in single level. When the equations are generalized linear models, such as logistic regressions, or even the structure is multilevel, the two estimators are not in the same scale, thus they are generally different (MacKinnon *et al.*, 1995; Iacobucci, 2012). The mediated effect $\hat{\alpha}\hat{\beta}$ is usually preferred as it is easier to compute and generalizes to more complicated models. Sobel (1982) has proposed the following formula to estimate the standard error and confidence interval for the mediated effect in models with normally distributed variables.

$$\hat{\sigma}_{\hat{\alpha}\hat{\beta}} = \sqrt{\hat{\alpha}^2\hat{\sigma}_{\hat{\beta}}^2 + \hat{\beta}^2\hat{\sigma}_{\hat{\alpha}}^2}, \quad (3.5)$$

where $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\beta^2$ are sampling variances of $\hat{\alpha}$ and $\hat{\beta}$. A 95% CI can be approximated by a Normal distribution and $\hat{\alpha}\hat{\beta} \pm 1.96\hat{\sigma}_{\hat{\alpha}\hat{\beta}}$. However, this solution relies on large sample approximation and the skewed distribution of the product causes a loss of power of the mediation test. MacKinnon *et al.* (2007) have proposed the bootstrapping method to provide an empirical sampling distribution of the mediated effect $\alpha\beta$ by resampling the original data and applying mediation analysis to each one. Zhang *et al.* (2009) show that the bootstrap and Bayesian posterior estimates are close in the case of a simple mediation model.

If the mediator M and the dependent variable are dichotomous, GLM-based mediation analysis can be applied,

$$E(Y) = g_Y^{-1}(\beta_1 + \tau X), \quad (3.6)$$

$$E(M) = g_M^{-1}(\beta_2 + \alpha X), \quad (3.7)$$

$$E(Y) = g_Y^{-1}(\beta_3 + \tau'X + \beta M), \quad (3.8)$$

where g_Y and g_M are link functions. For example, if Y follows a Bernoulli distribution, then g_Y can be a logit function. As proposed by Iacobucci (2012), one possible test for mediation is as follows,

1. For equation 3.6, if Y is continuous, fit a regression. If Y is categorical, fit it via a logistic regression.
2. For equation 3.7, if M is continuous, fit a regression. If M is categorical, fit it via a logistic regression. Collect the parameter estimate $\hat{\alpha}$, and its standard error $\hat{\sigma}_{\hat{\alpha}}$.
3. For equation 3.8, if Y is continuous, fit a regression. If Y is categorical, fit

it via a logistic regression. Collect the parameter estimate $\hat{\beta}$, and its standard error $\hat{\sigma}_{\hat{\beta}}$.

4. Standardize the estimated coefficients,

$$\begin{aligned} z_{\alpha} &= \hat{\alpha} / \hat{\sigma}_{\hat{\alpha}}, \\ z_{\beta} &= \hat{\beta} / \hat{\sigma}_{\hat{\beta}}. \end{aligned} \tag{3.9}$$

Compute their product, $z_{\alpha \times \beta} = z_{\alpha} z_{\beta}$, and the standard error: $\sqrt{z_{\alpha}^2 + z_{\beta}^2 + 1}$.

5. Perform the z -test,

$$z_{\text{mediation}} = \frac{z_{\alpha} z_{\beta}}{\sigma_{z_{\alpha \beta}}} = \frac{\frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} \times \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}}{\sqrt{z_{\alpha}^2 + z_{\beta}^2 + 1}}. \tag{3.10}$$

However, this z -test still shares the disadvantage of the test proposed by Sobel (1982), since the product of coefficients can't be assumed Normal or t distributed in real data. We also argue that this approach couldn't correctly test the mediated effect when the link function g_M is not an identity function, since in this case, Equation 3.4 doesn't hold anymore. Thus the product of two standardized coefficients doesn't represent the mediated effect and we can't consider it as a general solution to all link functions. Then, empirical-based modeling techniques such as bootstrap and the Bayesian approach receive more and more attention.

3.2.2 Bayesian Mediation Analysis of Single-Level Models

As mentioned above, one advantage of Bayesian inference is that it can estimate mediation effect in a very straightforward way. That is, conditioning on the observation and the priors of the unknown parameters, estimations of these parameters are based upon their posterior distributions.

To apply the Bayesian method, normal priors are often used for parameters in equations (3.6) to (3.8), that is

$$\begin{aligned}\beta_1 &\sim N(\mu_{\beta_1}, \sigma_1), \beta_2 \sim N(\mu_{\beta_2}, \sigma_2), \beta_3 \sim N(\mu_{\beta_3}, \sigma_3) \\ \tau &\sim N(\mu_\tau, \sigma_\tau), \tau' \sim N(\mu_{\tau'}, \sigma_{\tau'}) \\ \alpha &\sim N(\mu_\alpha, \sigma_\alpha), \beta \sim N(\mu_\beta, \sigma_\beta)\end{aligned}\tag{3.11}$$

Here, $\mu_{\beta_i}, \sigma_i, i = 1, 2, 3, \mu_\tau, \mu_{\tau'}, \mu_\alpha, \mu_\beta, \sigma_\tau, \sigma_{\tau'}, \sigma_\alpha$ and σ_β are hyperparameters, which are predefined reflecting prior information on these parameters.

If the equations are simple linear models, it is not difficult to derive the posterior distributions of the important parameters α, β, τ and τ' . However, in most GLM approaches, such as equations (3.6) - (3.8), MCMC methods need to be used to generate posterior samples. As discussed in Chapter 3, the **WinBUGS** or **JAGS** software can be used to implement MCMC methods to obtain the posterior draws. The JAGS code for the single-level mediation analysis is given in Appendix A.1.

Let $\alpha^{(t)}, \beta^{(t)}$ denote the t th posterior draw of these parameters for $t = 1, \dots, T$. Point estimates of these parameters can be obtained as sample means of these posterior draws. For example, the point estimate of α is given by

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^T \alpha^{(t)}.\tag{3.12}$$

The posterior variance of α is given by

$$Var(\alpha|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (\alpha^{(t)} - \hat{\alpha})^2.\tag{3.13}$$

For the estimation of the mediated effect, similar to conventional mediation analysis, there are two ways to estimate it. One is based on equation 3.6 and 3.8 to

compute the difference of the effects of the independent variable X before and after the mediation ($\tau - \tau'$). The other one is to compute the product of $\alpha\beta$. In Bayesian perspective, all these parameters are random which means these two terms don't follow the same distribution in general. Although, we have $E(\tau - \tau') = E(\alpha\beta)$. The product estimation is usually preferred since it is very straightforward and easier to compute in linear regression models and can be easily extended to generalized linear models. We use the product as a measure of the mediation effect for all possible combinations of X , M , and Y being continuous or categorical. For simple linear regression models, the point estimate of the mediated effect is,

$$\hat{\alpha}\beta = \frac{1}{T} \sum_{t=1}^T (\alpha^{(t)}\beta^{(t)}). \quad (3.14)$$

The posterior variance is,

$$Var(\alpha\beta|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (\alpha^{(t)}\beta^{(t)} - \hat{\alpha}\beta)^2. \quad (3.15)$$

For generalized linear models (Equations 3.6 – 3.8), instead of standardizing the coefficients, we propose to use linear approximation of g_M^{-1} to obtain the first order term of X and then use the product of the coefficient of this term and β to estimate and test the mediated effect.

Assuming g_M^{-1} is twice continuously differentiable, the linear approximation at point a is as follows,

$$g_M^{-1}(\beta_2 + \alpha X) \approx g_M^{-1}(a) + (g_M^{-1})'(a)(\beta_2 + \alpha X - a). \quad (3.16)$$

The mediated effect is then defined as,

$$z_{\alpha \times \beta} = (g_M^{-1})'(a)\alpha \times \beta. \quad (3.17)$$

The point estimate of the mediated effect is,

$$\hat{z}_{\alpha\times\beta} = \frac{1}{T} \sum_{t=1}^T (z_{\alpha\times\beta}^{(t)}). \quad (3.18)$$

where $z_{\alpha\times\beta}^{(t)} = (g_M^{-1})'(a)\alpha^{(t)} \times \beta^{(t)}$. The posterior variance is,

$$Var(z_{\alpha\times\beta}|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (z_{\alpha\times\beta}^{(t)} - \hat{z}_{\alpha\times\beta})^2. \quad (3.19)$$

The 95% credible interval of $\alpha\beta$ and $z_{\alpha\times\beta}$ is given by $[q_{0.025}^*, q_{0.975}^*]$, where $q_{0.025}^*$ and $q_{0.975}^*$ denote the 0.025 and 0.975 sample quantiles of the posterior draws of $\alpha\beta$ and $z_{\alpha\times\beta}$, respectively. The Bayesian hypothesis test (2.25) can be applied to these estimates.

3.3 Bayesian Mediation Analysis of Multilevel Models

As discussed in previous chapters, in many fields of research, the structure of observational or experimental data often has two levels which typically include individual (within-subject) and population/group (between-subject) variables. Individuals in the same group or data samples from the same individual are assumed to follow the same distribution which is different from that of other groups or individuals. Thus the single-level mediation model is not appropriate anymore since the assumption of independent observations is violated.

In multilevel mediation, the mediator and independent variable may both occur in lower level (within-subject variables), upper level (between-subject variables) or in different levels (Kenny *et al.*, 1998, 2003; Krull and MacKinnon, 1999, 2001). Estimation of multilevel models is much more complex than single-level models,

since the correlation among parameters in multilevel models might be considered and maximum likelihood usually involves many parameters that can cause a problem of convergence. Kenny *et al.* (2003) discussed these difficulties and introduced a method to estimate the covariance among random effects. By using a selection variable, Bauer *et al.* (2006) proposed a method to obtain consistent estimates of the variance components by simultaneously fitting the two mediation regression equations. All these methods are reviewed and an general approach is summarized in MacKinnon (2008).

As introduced before, hierarchical Bayesian methods has its unique advantages in analyzing multilevel or hierarchical mediation models. First, in the Bayesian framework, parameters are considered as random variables instead of fixed values which naturally handle the case of multilevel models. That is, the lower (subject-) level parameters are assumed to follow a distribution which involves the upper (population-) level parameters that automatically constructs the upper level model. Second, by using MCMC methods, estimation in multilevel modeling is relatively straightforward under the Bayesian framework. Almost any model with various distributions and any number of levels can be estimated simply by deriving the conditional posteriors or using other MCMC sampling techniques and software. Third, as mentioned earlier, for small data sets, Bayesian inference is relatively better than others, since it does not assume the symmetry and normality of estimates. Yuan and MacKinnon (2009) studied the multilevel mediation analysis from the Bayesian perspective, but that study was limited to continuous variables. In this section, we extend the multilevel model to even more complicated situation, that is the mediator

and dependent variable can be either continuous or categorical in either the same or different levels. The models are as follows (**JAGS** code is included in the Appendix A.2 for each of the three cases),

1) Mediator (M) and independent variable (X) in subject-level

Subject-level model:

$$E(M_{i,j}) = g_M^{-1}(\beta_{2,j} + \alpha_j X_{i,j}), \quad (3.20)$$

$$E(Y_{i,j}) = g_Y^{-1}(\beta_{3,j} + \tau'_j X_{i,j} + \beta_j M_{i,j}). \quad (3.21)$$

Population-level model:

$$\begin{aligned} \beta_{2,j} &= \sum_{k=1}^K Z_{j,k} \theta_{1,k} + \mu_{1,j}, \\ \alpha_j &= \sum_{k=1}^K Z_{j,k} \theta_{2,k} + \mu_{2,j}, \\ \beta_{3,j} &= \sum_{k=1}^K Z_{j,k} \theta_{3,k} + \mu_{3,j}, \\ \tau'_j &= \sum_{k=1}^K Z_{j,k} \theta_{4,k} + \mu_{4,j}, \\ \beta_j &= \sum_{k=1}^K Z_{j,k} \theta_{5,k} + \mu_{5,j}, \end{aligned} \quad (3.22)$$

where $i, i = 1, \dots, n$, indexes the number of observation, $j, j = 1, \dots, J$, indexes the subject or group and $k, k = 1, \dots, K$, indexes the between-subject covariates. In the subject-level model, g_Y, g_M are link functions (if linear regressions are used, they are replaced with identity functions and residuals are assumed to follow normal distributions). $\beta_{2,j}$ and $\beta_{3,j}$ are random intercepts. τ'_j, α_j and β_j are random slopes. All coefficients in this level are random in order to capture the heterogeneity in higher level model. In the population-level model, $\theta_{i,k}, i = 1, \dots, 5$, denotes the

coefficient for each between-subject covariate indexed by k . $Z_{j,k}$ is an element of Z , a $J \times K$ matrix of between-subject covariates. Here we assume the between-subject residuals $\mu_j = (\mu_{1,j}, \mu_{2,j}, \dots, \mu_{5,j})^T$, follow a multivariate normal distribution,

$$\mu_j \sim N(\mathbf{0}, \mathbf{\Sigma}). \quad (3.23)$$

where $\mathbf{0}$ is a vector of 0s, and $\mathbf{\Sigma}$ is a 5×5 covariance matrix.

In multilevel mediation, it is often of interest to investigate the average indirect effect in the population level. Kenny *et al.* (2003) showed that the total effect in a fully random, lower level mediated multilevel model is

$$c = E(\tau'_j) + E(\alpha_j \beta_j) + \sigma_{\alpha_j \beta_j}, \quad (3.24)$$

where $\sigma_{\alpha_j \beta_j}$ denotes the covariance between α_j and β_j . The average indirect effect is,

$$ab = E(\alpha_j \beta_j). \quad (3.25)$$

For the case of subject-level generalized linear models, similar to the situation in single-level mediation models, we propose to use the approximated product to measure the indirect effect,

$$z_{\alpha \times \beta} = E((g_M^{-1})'(a) \alpha_j \beta_j). \quad (3.26)$$

To estimate the multilevel mediation models in Bayesian perspective, priors need to be assigned to all parameters in the model including regression parameters ($\theta_{i,k}$, $i = 1, \dots, 5$ and the covariance matrix $\mathbf{\Sigma}$). Independent normal prior distributions with appropriate hyperparameters are often assigned on these regression

parameters, as introduced in previous sections. We can also use noninformative uniform priors to simplify the model,

$$p(\theta_{1,k}, \theta_{2,k}, \dots, \theta_{5,k}, k = 1, \dots, K) \propto \text{Unif}(-\infty, +\infty). \quad (3.27)$$

For the covariance matrix Σ , the inverse Wishart distribution is often used as the prior distribution, which has two parameters: the degree of freedom parameter ν and the scale matrix Ψ , a 5×5 matrix.

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu). \quad (3.28)$$

In Bayesian statistics, the inverse Wishart distribution is often used as the conjugate prior for the covariance matrix of a multivariate normal distribution. To represent the least informative prior knowledge, the degree of freedom is set to be as small as possible (i.e., $\nu = 5$), and a diagonal matrix with small values is assigned as the scale matrix. After specifying the prior distributions of these parameters, the hierarchical Bayesian generalized linear model can be easily fitted using MCMC sampling software such as **WinBUGS** and **JAGS**.

Let $\alpha = \sum_{k=1}^K Z_{j,k} \theta_{2,k}$, $\beta = \sum_{k=1}^K Z_{j,k} \theta_{5,k}$, denote the means of α_j and β_j . Let $\alpha^{(t)}$ and $\beta^{(t)}$, $t = 1, \dots, T$, denote the t th posterior draws of these means and $\sigma_{\alpha_j \beta_j}^{(t)}$ denote the t th draw of the covariance. It is very straightforward to obtain the inference of the average indirect effect. For the effect in 3.25, the t th posterior draw of the mediated effect is,

$$ab^{(t)} = \alpha^{(t)} \beta^{(t)}. \quad (3.29)$$

Then the posterior mean and variance of the average indirect effect are given by

$$\hat{ab} = \frac{1}{T} \sum_{t=1}^T (ab^{(t)}). \quad (3.30)$$

The posterior variance is,

$$Var(ab|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (ab^{(t)} - \hat{ab})^2. \quad (3.31)$$

For the effect in 3.26, the t th posterior draw of the mediated effect is,

$$z_{\alpha \times \beta}^{(t)} = (g_M^{-1})'(a) \alpha^{(t)} \beta^{(t)}. \quad (3.32)$$

Then the posterior mean and variance of the average indirect effect are given by

$$\hat{z}_{\alpha \times \beta} = \frac{1}{T} \sum_{t=1}^T (z_{\alpha \times \beta}^{(t)}). \quad (3.33)$$

The posterior variance is,

$$Var(z_{\alpha \times \beta}|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (z_{\alpha \times \beta}^{(t)} - \hat{z}_{\alpha \times \beta})^2. \quad (3.34)$$

2) Mediator (M) and independent variable (X) occur in population-level

Subject-level model:

$$E(Y_{i,j}) = g_Y^{-1} \left(\sum_{p=0}^P \beta_{2j}^p Z'_{i,p} \right). \quad (3.35)$$

Population-level model:

$$\begin{aligned} \beta_{2,j}^p &= \theta_{2,p} + \beta_p M_j + \tau_p' X_j + \mu_{2,j}^p, \\ E(M_j) &= g_M^{-1}(\theta_M + \alpha X_j). \end{aligned} \quad (3.36)$$

where $i, i = 1, \dots, n$, indexes the number of observation, $j, j = 1, \dots, J$, indexes the subject or group and $p, p = 1, \dots, P$, indexes the within-subject covariates. In the

subject-level model, g_Y is a link function. $Z'_{i,p}$ is an element of Z' , a $n \times P$ matrix of within-subject covariates. $\beta_{2,j}^p$ are random coefficients for within-subject covariates. If there is no within-subject covariate, then only intercepts are included in the model (β_{2j}^0). Note that, different from the multilevel model where the mediator (M) and independent variable (X) are both at subject-level, here, for each within-subject covariate $p, p = 1, \dots, P$, there is one population-level model. $\theta_{2,p}$ and θ_M are intercepts. τ'_p and β_p are coefficients for the mediator M and independent variable X corresponding to each within-subject covariate indexed by p . α is the slope in the regression of M on X . Here, we assume the between-subject residual corresponding to each within-subject covariate $\mu_{2,j}^p$, follow a normal distribution,

$$\mu_{2,j}^p \sim N(0, \sigma_p). \quad (3.37)$$

Similar to the previous mediation model, priors are assigned to all unknown parameters in the model including regression parameters including $\theta_{2,p}, \theta_M, \tau'_p, \beta_p$ and α . Independent normal prior distributions with appropriate hyperparameters are assigned on these parameters. The noninformative uniform prior is also a choice as introduced before.

We are interested in the indirect effect at the population level. As discussed before, if the equation 3.36 is a ordinary least squares regression, for each within-subject covariate indexed by p , the mediated effect is defined as,

$$(\alpha\beta)_p = \alpha\beta^p. \quad (3.38)$$

Let $\alpha^{(t)}$ and $\beta^{p(t)}$ denote the t th draw of these parameters. Then the t th draw of

the mediated effect is,

$$(\alpha\beta)_p^{(t)} = \alpha^{(t)}\beta^{p(t)}. \quad (3.39)$$

The posterior mean and variance of the mediated effect are given by

$$(\hat{\alpha}\hat{\beta})_p = \frac{1}{T} \sum_{t=1}^T ((\alpha\beta)_p^{(t)}). \quad (3.40)$$

The posterior variance is,

$$Var((\alpha\beta)_p|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T ((\alpha\beta)_p^{(t)} - (\hat{\alpha}\hat{\beta})_p)^2. \quad (3.41)$$

If the equation 3.36 is a generalized linear model, we again use the approximated product to estimate and test the mediated effect.

$$z_{\alpha \times \beta^p}^{(t)} = (g_M^{-1})'(a)\alpha^{(t)}\beta^{p(t)}, \quad (3.42)$$

The point estimate of the mediated effect is,

$$\hat{z}_{\alpha \times \beta^p} = \frac{1}{T} \sum_{t=1}^T (z_{\alpha \times \beta^p}^{(t)}). \quad (3.43)$$

The posterior variance is,

$$Var(z_{\alpha \times \beta^p}|\text{data}) = \frac{1}{T-1} \sum_{t=1}^T (z_{\alpha \times \beta^p}^{(t)} - \hat{z}_{\alpha \times \beta^p})^2. \quad (3.44)$$

3) Mediator (M) and independent variable (X) at different level

When the mediator is at the subject-level and the independent variable is at the population-level, it is straightforward to use hierarchical Bayesian ANOVA models introduced in the previous chapter to estimate parameters. That is,

Subject-level model:

$$\begin{aligned} E(Y_{i,j}) &= g_Y^{-1}(\beta_{0,j} + \beta_j M_{i,j}), \\ E(M_{i,j}) &= g_M^{-1}(\gamma_{0,j}), \end{aligned} \quad (3.45)$$

Population-level model:

$$\begin{aligned}
\beta_{0,j} &= \theta_0 + \alpha_0 X_j + \mu_{0,j}, \\
\beta_j &= \theta_1 + \alpha_1 X_j + \mu_{1,j}, \\
\gamma_{0,j} &= \theta_2 + \alpha X_j + \mu_{2,j}.
\end{aligned} \tag{3.46}$$

where $i, i = 1, \dots, n$, indexes the number of observation, $j, j = 1, \dots, J$, indexes the subject or group. In the subject-level model, g_Y and g_M are link functions. The within-subject covariates include the intercept $\beta_{0,j}, \gamma_{0,j}$ and the mediator $M_{i,j}$. β_j is a random coefficient for the mediator. Note that, different from the previous multilevel models where the mediator (M) and independent variable (X) are both at the same level, here, since they are at different levels (regular regression doesn't work), we propose to use a second hierarchical regression of M on X (Equation 3.45), where the intercept $\gamma_{0,j}$ is the only within-subject variable. In the population-level model, θ_0, θ_1 and θ_2 are intercepts. α_0 is the slope in the regression of the within level intercept $\beta_{0,j}$ on X which represents the main effect of X (similar to τ' in Equation 3.21). α_1 is the slope in the regression of the within level coefficient β_j on X which represents the interaction effect of the mediator M and the independent variable X . α represents the effect of X on M in a hierarchical way. We assume the between-subject residual $\mu_{0,j}, \mu_{1,j}$ and $\mu_{2,j}$ follow Normal distributions. Compared with Equations 2.1, 2.2 and 2.4a–d, we can see that the multilevel mediation model we introduced in Equation 3.45 and 3.46 is exactly a hierarchical Bayesian ANOVA model. Thus we can easily use the R package **BANOVA** to estimate all parameters.

Let β denote the expectation of β_j , that is, $\beta = E(\beta_j) = \theta_1 + \alpha_1 X_j$. The

average mediated effect is defined as,

$$ab = E(\alpha\beta_j) = \alpha\beta. \quad (3.47)$$

For the case of subject-level generalized linear models, similar to the situation in previous mediation models, we propose to use the approximated product to measure the indirect effect,

$$z_{\alpha\times\beta} = E((g_M^{-1})'(a)\alpha\beta_j) = (g_M^{-1})'(a)\alpha\beta. \quad (3.48)$$

Then following the methods introduced in previous models (Such as, Equations 3.29–3.34), we can compute several statistics of interest and test the mediated effect based on posterior samples.

3.4 Applications

3.4.1 Application 1: Single-Level Mediation

In this section, we use the Morse *et al.* (1994) data that were reanalyzed by Kenny *et al.* (1998). The independent variable, X is Treatment, a manipulated variable with 42.2% Controls and 57.8% Treateds, the dependent variable, Y is Days Housed, and the mediator or M is Housing Contacts. The causal relation is described as follows: The variable Treatment is presumed to cause Housing Contacts, which in turn is presumed to cause Days Housed (Morse *et al.*, 1994). There are a total of 109 observations. The means and standard deviations are presented in Table 3.1.

Variable	Mean	Standard Deviation
Treatment	.422	.496
Days Housed	15.552	13.107
Housing Contacts	3.462	3.843

Table 3.1: Descriptive statistics of housing data

The results following standard procedure introduced by Baron and Kenny (1986) are summarized in Table 3.2. From the result, we can see that the indirect effect of Treatment on Days Housed ($\alpha\beta$) is equal to 2.560, and the direct effect is equal to 3.998. The Sobel standard error (3.5) is equal to 1.157, which makes the Z test of the indirect effect equal to 2.213 ($p = .027$). Because the Sobel test is statistically significant, it is concluded that the indirect effect is significantly different from zero.

Path	Estimate	95% CI	p-value
τ	6.558	1.654 to 11.462	0.009
α	1.831	.389 to 3.274	0.013
β	1.398	.801 to 1.995	< 0.001
τ'	3.998	-.625 to 8.621	0.089

Table 3.2: Conventional mediation analysis results of housing data

We began the single-level Bayesian mediation analysis using simple linear regression, that is, in equations 3.6 to 3.8, the link functions are identity functions,

$$Y_i = \beta_1 + \tau X_i + \epsilon_i^1, \quad (3.49)$$

$$M_i = \beta_2 + \alpha X_i + \epsilon_i^2, \quad (3.50)$$

$$Y_i = \beta_3 + \tau' X_i + \beta M + \epsilon_i^3. \quad (3.51)$$

where independent normal prior information 3.11 are used for all parameters. The

hyper-parameters are set as,

$$\begin{aligned}
 \mu_{\beta_1} &= \mu_{\beta_2} = \mu_{\beta_3} \\
 &= \mu_{\tau} = \mu'_{\tau} = \mu_{\alpha} = \mu_{\beta} = 0, \\
 \sigma_1 &= \sigma_2 = \sigma_3 = \sigma_{\tau} \\
 &= \sigma_{\tau'} = \sigma_{\alpha} = \sigma_{\beta} = 10^{-6}.
 \end{aligned} \tag{3.52}$$

The residuals ϵ^1, ϵ^2 and ϵ^3 are assumed to follow Normal distribution,

$$\epsilon^i \sim Normal(0, \sigma_i^{-2}), \quad i = 1, 2, 3. \tag{3.53}$$

where priors for $\sigma_i, i = 1, 2, 3$, are assumed to follow *Gamma*(1, 1).

We use 5,000 iteration to burn-in and collected 5,000 posterior draws thinned by a factor 5 to make inference. By using either visual or formal diagnostics introduced before, we make sure the MCMC chains of all parameters converges well after burn-in period. The estimation results are summarized in Table 3.3.

Path	Estimate	95% CI	Bayesian p-value
τ	6.56	1.603 to 11.45	0.009
α	1.84	.354 to 3.30	0.015
β	1.40	.797 to 2.00	< 0.001
τ'	3.98	-.702 to 8.68	0.097

Table 3.3: Bayesian mediation analysis results of housing data

As you can see from the above results, Bayesian estimations are very close to conventional estimations. The posterior distribution of the mediated effect $\alpha\beta$ is displayed in Figure 3.2. The posterior mean and posterior standard error of $\alpha\beta$ are 2.58 and 1.203, respectively. The 95% credible interval for the mediated

effect is $[0.478, 5.17]$ which doesn't include 0. Thus the mediated effect is considered as significant. Moreover, Bayesian inference based on the posterior distribution automatically take into account of the skewed distribution of the mediated effect shown in Figure 3.2.

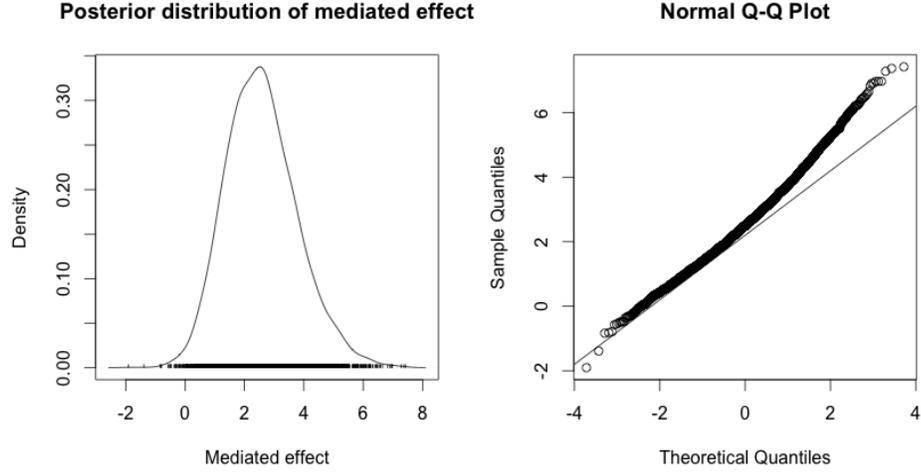


Figure 3.2: Posterior distribution of the mediated effect (left panel) and the corresponding normal quantile-quantile plot (right panel) for the Bayesian single-level mediation analysis of the housing data.

Next, we assume the dependent variable Y and the mediator M to follow Poisson distribution, since both of them are counting variables. The independent variable X remains the same. The new model is as follows,

$$\begin{aligned}
 Y_i &\sim \text{Poisson}(p_{1i}^Y), p_{1i}^Y = \exp(\beta_1 + \tau X_i), \\
 M_i &\sim \text{Poisson}(p_i^M), p_i^M = \exp(\beta_2 + \alpha X_i), \\
 Y_i &\sim \text{Poisson}(p_{2i}^Y), p_{2i}^Y = \exp(\beta_3 + \tau' X_i + \beta M_i).
 \end{aligned}
 \tag{3.54}$$

To conduct Bayesian analysis, the same independent normal priors (3.52) with the previous example are assigned to parameters in Equation 3.54. We use 10,000 it-

eration to burn-in and collected 2,000 posterior draws thinned by a factor 5 to make inference. Convergence of the MCMC chains of all parameters are ensured by diagnostics introduced before. The results are summarized in Table 3.4.

Path	Conventional Analysis			Bayesian Analysis		
	Estimate	95% CI	p-value	Estimate	95% CI	p-value
τ	0.413	0.318 to 0.509	< 2e - 16	0.413	0.315 to 0.513	< 5e - 04
α	0.501	0.298 to 0.704	1.3e - 06	0.500	0.296 to 0.711	< 5e - 04
β	0.069	0.059 to 0.079	< 2e - 16	0.069	0.059 to 0.079	< 5e - 04
τ'	0.286	0.189 to 0.383	6.97e - 09	0.285	0.191 to 0.379	< 5e - 04

Table 3.4: Comparison of conventional and Bayesian mediation analysis of housing data with dichotomous variables

Using the test proposed by Iacobucci (2012) (3.10), the mediated effect is estimated as $z_{mediation} = 4.568$, which makes the z -test p-value equal to $4.90e-06$. In this situation, the mediated effect is strongly significant. Compared with the conventional z -test, Bayesian methods estimate the mediated effect (3.18), $\hat{z}_{\alpha \times \beta} = 0.035$ which yields the Bayesian p-value less than $5e-04$. Note that, here we use the linear approximation at 0 of the log link function ($g_M = \ln, g_M^{-1}(\beta_2 + \alpha X_i) = \exp(\beta_2 + \alpha X_i) \approx 1 + 1 \cdot (\beta_2 + \alpha X_i)$). Bayesian inference successfully captures the slight non-normality of the mediated effect shown in Figure 3.3 and produces a reasonable test of the indirect effect.

3.4.2 Application 2: Multilevel Mediation

In this section, we revisit the housing data analyzed in the previous section. The data was originally designed as a between-subject data which has the person

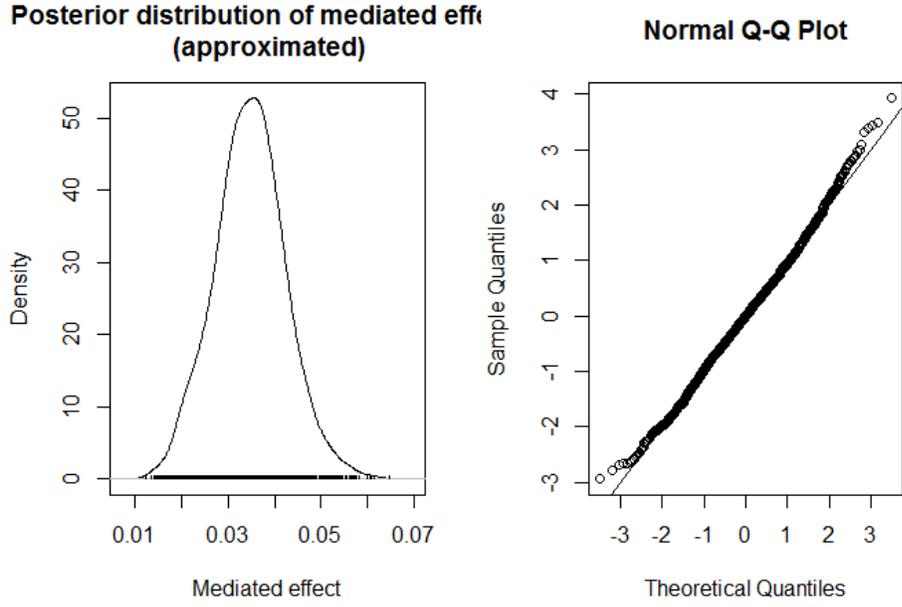


Figure 3.3: Posterior distribution of the mediated effect(left panel) and the corresponding normal quantile-quantile plot (right panel) for the Bayesian single-level mediation analysis of the housing data.

id information. Thus, we can apply the Bayesian multilevel mediation model in which both the mediator and independent variable are between-subject variables (see 3.36). The model is as follows,

Subject-level model:

$$\begin{aligned}
 Y_{i,j} &\sim \text{Poisson}(p_{i,j}^Y), \\
 \log(p_{i,j}^Y) &= \beta_{2,j}.
 \end{aligned} \tag{3.55}$$

Population-level model:

$$\begin{aligned}
 \beta_{2,j} &= \theta_2 + \beta M_j + \tau' X_j + \mu_{2,j}, \\
 M_j &\sim \text{Poisson}(p_j^M) \\
 \log(p_j^M) &= \theta_M + \alpha X_j.
 \end{aligned} \tag{3.56}$$

Note that, in this case, only the intercept $\beta_{2,j}$ is included in the subject-level model. The population-level residual is assumed to follow a normal distribution, $\mu_{2,j} \sim Normal(0, \sigma_2^{-2})$, where the prior for σ_2 is set as $Gamma(1, 1)$. Priors for other parameters in population-level are assumed to follow independent normal distributions, $\theta_2, \beta, \tau', \theta_M, \alpha \sim Normal(0, 10^{-6})$. We use 5,000 iterations as burn-in period to make sure the convergence (result not shown here) and collect 5,000 posterior samples thinned by a factor 5 to make inference. The result of the Bayesian analysis are reported in Table 3.5.

Effect	Estimate	95% CI	Bayesian p-value
α	0.502	0.300 to 0.708	< 0.001
β	0.206	0.100 to 0.316	< 0.001
τ'	0.774	-0.042 to 1.663	0.063

Table 3.5: Bayesian multilevel mediation analysis of housing data

We are interested in the test of the indirect effect in population level. By using the linear approximation at 0 of the log link function and the estimation introduced in Equation 3.42, the trace plot and density estimation of $z_{\alpha \times \beta}$ is shown in Figure 3.4, with mean equal to 0.103 (95% CI [0.043, 0.178]). Since the 95% credible interval doesn't contain 0, the result confirms the previous result that the mediated effect is significant in the model with dichotomous dependent and mediators.

3.5 Conclusion

In this chapter, we propose a Bayesian approach to mediation analysis in both single-level and multilevel models including either continuous or dichotomous

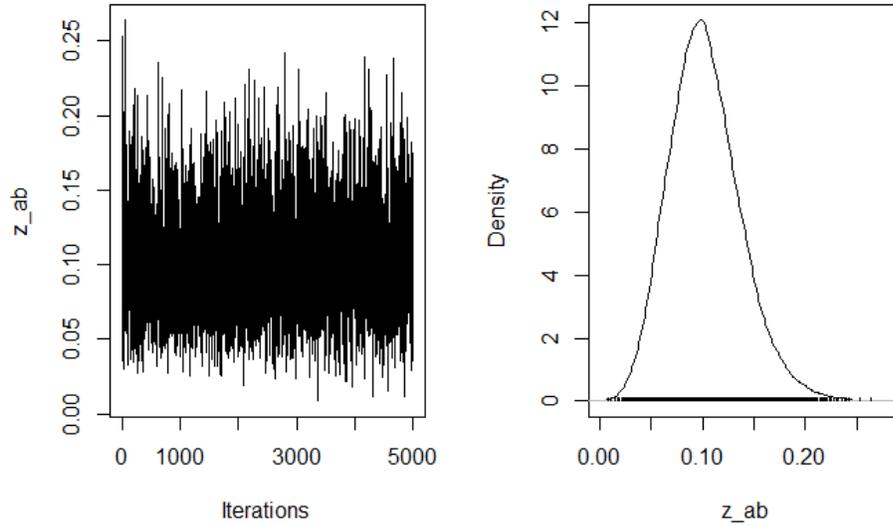


Figure 3.4: Trace and density plot of the posterior distribution of the mediated effect $z_{\alpha \times \beta}$ in between-subject level multilevel mediation model.

outcomes and mediators. The Bayesian approach is especially useful for single-level mediation analysis with generalized linear regressions and multilevel mediation analysis, which are not easily to handle through a conventional frequentist approach. Bayesian mediation analysis has many advantages including the natural way of setting up the model and computationally sampling the statistics of interest. By utilizing well-developed MCMC methods, along with several powerful software, such as **WinBUGS** and **JAGS**, researchers can easily build up any complex hierarchical models with least constraint. Compared with the frequentist approach, Bayesian inference based on posterior samples generated by MCMC methods does not depend on large-sample approximation. The inference is exact for small samples. This feature makes the Bayesian approach especially appealing for studies with small sample

sizes, which is the case in many experimental studies. Furthermore, Bayesian methods naturally incorporate prior research, allowing for the accumulation of knowledge of mediated effects across research studies. However, this might also be an issue in Bayesian inference. When the prior is inappropriately specified, estimates based on Bayesian mediation models can be biased. Thus, it is very critical to choose an appropriate prior to increase the accuracy of the estimation.

Chapter 4: Future Work

In this dissertation, we introduced hierarchical Bayesian ANOVA models with an accompanying R package to analyze experimental data including within- and between- subject variables. The response variable follows a wide variety of distributions 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. The **BANOVA** package simulate the analysis under reasonable assumptions on the distribution of the data and its hierarchical structure. It offers several user-friendly routines for the applied researcher and makes the life easier. The package can also be extended with many other models, for example using different distributions of the dependent variable. As discussed before, 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 to the R package, the web application for hierarchical Bayesian ANOVA is also introduced. Currently, there are six models included in the web. Users can

easily use these models by just uploading a data set without installing any software. In the future, more models included in the R package will be embedded into the web application, so that researchers who are not familiar with R programming language can easily access the models we developed.

We also introduced Bayesian mediation models in both single-level and multi-level. The Bayesian approach shows its advantages in testing mediated effects using generalize linear and multilevel mediation models. Bayesian approach is conceptually natural and easy to implement by using well-developed MCMC methods, which enable researchers to easily fit very complex generalized linear and hierarchical models. Especially when a continuous mediator is included in a hierarchical model, the computations can be done with the R package **BANOVA**. As discussed before, the Bayesian inference does not depend on large-sample approximation. It is exact for any size of samples. However, one important topic we have not covered in this dissertation is the mediation models involving multiple continuous or categorical mediators. If these mediators and the independent variable are at the same level, we can use multiple regressions for these mediators to deal with this case. However, if these mediators and the independent variable are at different levels, the hierarchical Bayesian models introduced before need to be modified to correctly represent the mediated effect. It would be a reasonable future research topic in Bayesian mediation analysis.

Chapter A: **JAGS** Code for Bayesian Mediation Analysis

A.1 **JAGS** Code for Single-level Mediation

```
model{  
  
  # Single-level mediation model, in which the dependent  
  # variable and mediator assumed to follow the Normal distribution.  
  # y[i], m[i], and x[i] denote data vectors of dependent  
  # variable, mediating variable and independent variable  
  # , respectively. N is the number of observations,  
  # ydup[i] is the duplicate of y[i].  
  for (i in 1:N){  
  
    ydup[i] ~dnorm(mean.ydup[i], prec.ydup)  
  
    mean.ydup[i] <- beta1 + tau*x[i]  
  
    m[i] ~dnorm(mean.m[i], prec.m)  
  
    mean.m[i] <- beta2 + alpha*x[i]  
  
    y[i] ~dnorm(mean.y[i], prec.y)  
  
    mean.y[i] <- beta3 + beta*m[i] + tau.prime*x[i]  
  
  }  
  
  # specify priors  
  
  beta1 ~dnorm(0, 0.000001)
```

```

beta2 ~dnorm(0, 0.000001)
beta3 ~dnorm(0, 0.000001)
alpha ~dnorm(0, 0.000001)
beta ~dnorm(0, 0.000001)
tau ~dnorm(0, 0.000001)
tau.prime ~dnorm(0, 0.000001)
prec.y ~dgamma(0.001, 0.001)
prec.ydup ~dgamma(0.001, 0.001)
prec.m ~dgamma(0.001, 0.001)
# specify the mediated effect
tau_m <- alpha*beta
}

model{
# Single-level mediation model, in which the dependent
# variable and mediator assumed to follow the Poisson distribution.
# y[i], m[i], and x[i] denote data vectors of dependent
# variable, mediating variable and independent variable
# , respectively. N is the number of observations,
# ydup[i] is the duplicate of y[i].
for (i in 1:N){
  ydup[i] ~ dpois(mean.ydup[i])
  log(mean.ydup[i]) <- beta1 + tau*x[i]

  m[i] ~ dpois(mean.m[i])
  log(mean.m[i]) <- beta2 + alpha*x[i]

  y[i] ~ dpois(mean.y[i])

```

```

    log(mean.y[i]) <- beta3 + beta*m[i] + tau.prime*x[i]
  }

# specify priors

beta1 ~dnorm(0, 0.000001)
beta2 ~dnorm(0, 0.000001)
beta3 ~dnorm(0, 0.000001)
tau ~dnorm(0, 0.000001)
alpha ~dnorm(0, 0.000001)
beta ~dnorm(0, 0.000001)
tau.prime ~dnorm(0, 0.000001)

# specify mediated effect

tau_m <- alpha*beta
}

```

A.2 JAGS Code for Multilevel Mediation

```

model
{
  # specify the multilevel model, in which the dependent variable
  # assumed to follow the Normal distribution. The mediator and the independent variable
  # are both at the lower level. Proposed by Yuan and MacKinnon (2009)
  # N1 and N2 are the number of first-level units and
  # second-level units, respectively.

  for (j in 1:N2){
    for (i in 1:N1){
      # Specify the first-level models

      ydup[j,i] ~dnorm(mean.direc[j,i], prec.direc)
    }
  }
}

```

```

mean.direc[j,i] <- Beta1[j] + Tau.direc[j] * x[j,i]

m[j,i] ~dnorm(mean.m[j,i], prec.m)

mean.m[j,i] <- Beta2[j] + AlphaBeta[j,1] * x[j,i]

y[j,i] ~ dnorm(mean.y[j,i], prec.y)

mean.y[j,i] <- Beta3[j] + AlphaBeta[j,2] * m[j,i] + Tau.p[j] * x[j,i]
}

# Specify the second-level models
Beta1[j] ~ dnorm(beta1, prec.beta1)
Tau.direc[j] ~ dnorm(tau.direc, prec.taudirec)
Beta2[j] ~ dnorm(beta2, prec.beta2)
Beta3[j] ~ dnorm(beta3, prec.beta3)
Tau.p[j] ~ dnorm(tau.p, prec.taup)

# bivariate normal distribution for alpha and beta
AlphaBeta[j, 1:2] ~dmnorm(alphabeta[], prec.ab[,])
}

alphabeta[1:2] ~ dmnorm(mean[], prec[,])

# ...

# some code are skipped, please refer to the original paper (Yuan and MacKinnon,2009)

# for more details.

# ...

# Wishart prior for the covariance matrix
prec.ab[1:2,1:2] ~ dwish(Omega[,], 2)

# ...

# some code are skipped, please refer to the original paper (Yuan and MacKinnon,2009)

```

```

# for more details.

# ...

# specify the mediated effect
tau_m <- alphabeta[1]*alphabeta[2]
}

model{

# specify the multilevel model, in which the dependent variable
# and the mediator assumed to follow the Poisson distribution, and between-subject only
# N and M are the number of first-level units and second-level units, respectively.
# specify the first-level model
for (i in 1:N){
  y[i] ~dpois(py[i])
  log(py[i]) <- beta2[id[i]]
}

# specify the second-level models
for (i in 1:M){
  m[i] ~dpois(pm[i])
  log(pm[i]) <- thetam + alpha*x[i]

  beta2[i] ~dnorm(mu.beta2[i], tau.beta2)
  mu.beta2[i] <- theta2 + beta*m[i] + tau.prime*x[i]
}

thetam ~dnorm(0, 0.000001)
theta2 ~dnorm(0, 0.000001)
alpha ~dnorm(0, 0.000001)
beta ~dnorm(0, 0.000001)
tau.prime ~dnorm(0, 0.000001)

```

```
tau.beta2 <- pow(sigma.beta2, -2)
sigma.beta2 ~dgamma(1,1)
# specify the mediated effect
tau_m <- alpha*beta
}
```

Chapter B: **BANOVA** User Manual

Package ‘BANOVA’

June 10, 2014

Type Package

Title BANOVA: An R-Package for Hierarchical Bayesian ANOVA

Version 0.1

Date 2014-05-11

Author Chen Dong, Michel Wedel

SystemRequirements jags (see <http://mcmc-jags.sourceforge.net>)

Maintainer Chen Dong <chendong@umd.edu>

Depends R (>= 3.0.1)

Imports runjags, coda

Description BANOVA covers several Bayesian Analysis of Variance (BANOVA) models used in analysis of experimental designs in which both within- and between- subjects factors are manipulated. They can be applied to data that are common in the behavioral sciences and related fields. The package includes: Hierarchical Bayes ANOVA models with normal response, t response, Binomial(Bernoulli) response, Poisson response, ordered multinomial response and multinomial response variables. All models accommodate unobserved heterogeneity by including a normal distribution of the parameters across individuals. Outputs of the package include tables of sums of squares, effect sizes and p-values, and tables of means, which are easily interpretable for behavioral researchers. BANOVA uses JAGS as a computational platform.

License GPL (>= 2)

R topics documented:

BANOVA-package	95
BAnova	96
BANOVA.Bern	97
BANOVA.Bin	99
BANOVA.Multinomial	101
BANOVA.Normal	103
BANOVA.ordMultinomial	104

BANOVA.Poisson	106
BANOVA.T	108
bernlogtime	110
bpndata	111
choicedata	112
colorad	114
colorad2	115
conv.diag	116
goalstudy	117
ipadstudy	119
table.means	120
table.pvalues	121
trace.plot	122

 BANOVA-package

BANOVA: An R-Package for Hierarchical Bayesian ANOVA

Description

This package include several hierarchical Bayes Analysis of Variance models. These models are suited for the analysis of experimental designs in which both within- and between- subjects factors are manipulated, and account for a wide variety of distributions of the dependent variable.

Details

Package: BANOVA
 Type: Package
 Version: 0.1
 Date: 2014-05-11
 License: GPL (>= 2)

Model:

$$E(y_i) = g^{-1}(\eta_i)$$

where $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of data response i . The within-subject factors and their interactions are indexed by p ($p = 1, 2, \dots, P$). Each index p represents a batch of J_p coefficients: $\beta_{j,s}^p, j = 1, \dots, J_p; s = 1, \dots, S$ indexes subjects. Note that if the subject-level covariate is continuous, $J_p = 1$, so that ANCOVA models are also accommodated (relaxing their "constant slope" assumption).

The population-level model allows for heterogeneity among subjects, because the subject-level coefficients $\beta_{j,s}^p$ are assumed to follow a multivariate normal distribution. The between-subject factors and their interactions are indexed by q , ($q = 1, 2, \dots, Q$), $q = 0$ denotes the constant term. The population-level ANOVA can be written as:

$$\beta_{j,s}^p = \sum_{q=0}^Q \theta_{j,k_s^q}^{pq} + \delta_{j,s}^p$$

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:

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

where $Z_{s,k}$ is an element of Z , a $S \times Q$ matrix of covariates. $\theta_{j,k}^p$ is a hyperparameter which captures the effects of between-subject factor q on the parameter $\beta_{j,s}^p$ of within-subject factor p . The error $\delta_{j,s}^p$ is assumed to be normal: $\delta_{j,s}^p \sim N(0, \sigma_p^{-2})$. Proper, but diffuse priors are assumed: $\theta_{j,k}^p \sim N(0, s)$, and $\sigma_p \sim \text{Gamma}(a, b)$, where s, a, b are hyper-parameters. The default setting is $s = 10^{-4}, a = 1, b = 1$.

Author(s)

Chen Dong; Michel Wedel

Maintainer: Chen Dong <chendong@umd.edu>

References

McCullagh, P., Nelder, JA. (1989) *Generalized linear models*, New York, NY: Chapman and Hall.

Gelman, A. (2005) *Analysis of variance-why it is more important than ever*, Ann. Statist. Vol. 33, No. 1, 1-53.

Rossi, P., Allenby, G., McCulloch, R. (2005) *Bayesian Statistics and Marketing*, John Wiley and Sons.

Gill, J. (2007) *Bayesian Methods for the Social and Behavioral Sciences*, Chapman and Hall, Second Edition.

Gelman, A., Carlin, J., Stern, H. and Dunson, D. (2013). *Bayesian Data Analysis*, London: Chapman & Hall.

BANova

Function to print the table of effect sizes

Description

The analysis of variance is performed at level 2 equation of the Bayesian ANOVA see [BANOVA-package](#). This makes it possible to capture the effects of level-2 variables on the heterogeneity distribution of subjects, and compute sums of squares and effect sizes.

Usage

BANova(x)

Arguments

x the object from BANOVA.*

Details

Measures of effect size in regression are measures of the degree of association between an effect (e.g., a main effect, an interaction, a linear contrast) and the dependent variable. They can be considered as the correlation between a categorical factor(effect) and the dependent variable. They are usually interpreted as the proportion of variance in the dependent variable that is attributable to each effect. In the package, Eta squared is calculated and displayed. It is defined as follows,

$$\eta^2 = \frac{(SS_{effect})}{(SS_{total})}$$

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.

If the design is not balanced, type II(if interactions are not included in the model) and type III sum of squares are computed.

References

Fox, J. (2008) *Applied Regression Analysis and Generalized Linear Models*, Second Edition. Sage.

Fox, J. and Weisberg, S. (2011) *An R Companion to Applied Regression*, Second Edition, Sage.

Examples

```
## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
BANova(res)

## End(Not run)
```

BANOVA.Bern

Estimation of BANOVA with a Bernoulli dependent variable

Description

BANOVA.Bern implements a Bayesian ANOVA for binary dependent variable, using a logit link and a normal heterogeneity distribution.

Usage

```
BANOVA.Bern(l1_formula = "NA", l2_formula = "NA", data,
  id, burnin = 500, sample = 1000, thin = 1, jags = findjags())

## S3 method for class BANOVA.Bern
summary(object, ...)
```

```
## S3 method for class BANOVA.Bern
predict(object, newdata = NULL, ...)
## S3 method for class BANOVA.Bern
print(x, ...)
```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data.frame in long format including all features in level 1 and level 2(covariates and categorical factors) and responses
id	subject ID of each response unit
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.Bern (returned by BANOVA.Bern)
newdata	test data, either a matrix, vector or a data.frame. It must have the same format with the original data (the same number of features and the same data classes)
x	object of class BANOVA.Bern (returned by BANOVA.Bern)
...	additional arguments, currently ignored

Details

Level 1 model:

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

where $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of data record i . see [BANOVA-package](#)

Value

BANOVA.Bern returns an object of class "BANOVA.Bern". The returned object is a list containing:

anova.table	table of effect sizes BANova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
data	original data.frame
mf1	model.frame of level 1
mf2	model.frame of level 2
JAGSmodel	JAGS model

Examples

```
## Not run:
# use the data in bernlogtime
data(bernlogtime)
# model with the dependent variable : response
res <- BANOVA.Bern(response~typical, ~blur + color, bernlogtime,
bernlogtime$subject)
summary(res)
table.means(res)

## End(Not run)
```

BANOVA.Bin

Estimation of BANOVA with a Binomial dependent variable

Description

BANOVA.Bin implements a Hierarchical Bayesian ANOVA for a binomial response variable using a logit link and a normal heterogeneity distribution.

Usage

```
BANOVA.Bin(l1_formula = "NA", l2_formula = "NA", data,
  id, num_trails, burnin = 500, sample = 1000, thin = 1, jags = findjags())
## S3 method for class BANOVA.Bin
summary(object, ...)
## S3 method for class BANOVA.Bin
predict(object, newdata = NULL, ...)
## S3 method for class BANOVA.Bin
print(x, ...)
```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data.frame in long format including all features in level 1 and level 2(covariates and categorical factors) and responses
id	subject ID of each response unit
num_trails	the number of trails of each observation(=1, if it is bernoulli), the type is forced to be 'integer'
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned

jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.Bin (returned by BANOVA.Bin)
newdata	test data, either a matrix, vector or a data frame. It must have the same format with the original data (the same column number)
x	object of class BANOVA.Bin (returned by BANOVA.Bin)
...	additional arguments, currently ignored

Details

Level 1 model:

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

where n_{trails} is the binomial total for each record i , $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of response i . see [BANOVA-package](#)

Value

BANOVA.Bin returns an object of class "BANOVA.Bin". The returned object is a list containing:

anova.table	table of effect sizes BANova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
data	original data.frame
mf1	model.frame of level 1
mf2	model.frame of level 2
JAGSmodel	JAGS model

Examples

```
## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
summary(res)
table.means(res)
# use the the first row of data to predict,
# only the in-model variables data will be used, the others will be ignored
predict(res, c(1, 0, 8, 2, 1, 0.03400759))

## End(Not run)
```

BANOVA.Multinomial *Estimation of BANOVA with a Multinomial dependent variable*

Description

BANOVA.Multinomial implements a Hierarchical Bayesian ANOVA for multinomial response variable using a logit link and a normal heterogeneity distribution.

Usage

```
BANOVA.Multinomial(l1_formula = "NA", l2_formula = "NA",
  dataX, dataZ, y, id, burnin = 500, sample = 1000, thin = 1, jags = findjags())
## S3 method for class BANOVA.Multinomial
summary(object, ...)
## S3 method for class BANOVA.Multinomial
predict(object, Xsamples = NULL, Zsamples = NULL, ...)
## S3 method for class BANOVA.Multinomial
print(x, ...)
```

Arguments

l1_formula	formula for level 1 e.g. '~X1+X2', response variable must not be included
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
dataX	a list of data frames(each corresponds to the choice set of each observation) that includes all covariates and factors
dataZ	a data frame(long format) that includes all level 2 covariates and factors
y	choice responses, 1,2,3...
id	subject id
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.Multinomial(returned by BANOVA.Multinomial)
Xsamples	new data samples in level one, must be a list(the same format with the training data), numeric variables must be mean centered.
Zsamples	new data samples in level two(the same format with the training data), numeric variables must be mean centered.
x	object of class BANOVA.Multinomial (returned by BANOVA.Multinomial)
...	additional arguments,currently ignored

Details

Level 1 model:

$$P(y_i = k) = \frac{\exp(\eta_{ik})}{\sum_{k=1}^K \exp(\eta_{ik})}$$

where $\eta_{ik} = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^{k,p} \beta_{j,s_i}^p$, s_i is the subject id of response i , see [BANOVA-package](#). $X_{i,j}^{k,p}$ is the design matrix corresponding to each class k ($k = 1, \dots, K$) of y_i . The first level of the response is the base level, thus the intercept corresponding to this level will not be included.

Value

BANOVA.Multinomial returns an object of class "BANOVA.Multinomial". The returned object is a list containing:

anova.table	table of effect sizes BANova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
dataX	original dataX
dataZ	original dataZ
mf1	model.frame of level 1
mf2	model.frame of level 2
n_categories	the number of categories of the response
JAGSmodel	JAGS model

Examples

```
## Not run:
# see choicedata
data(choicedata)
# generate dataX(convert the within-subject variables to a list)
dataX <- list()
for (i in 1:nrow(choicedata)){
  logP <- as.numeric(log(choicedata[i,3:8]))
  # all numeric variables must be mean centered
  dataX[[i]] <- as.data.frame(logP) - mean(logP)
}

dataZ <- choicedata[,9:13]
res <- BANOVA.Multinomial(~ logP, ~ college, dataX, dataZ, choicedata$choice, choicedata$hhid)

## End(Not run)
```

BANOVA.Normal

*Estimation of BANOVA with a normally distributed dependent variable***Description**

BANOVA.Normal implements a Hierarchical Bayesian ANOVA for linear models with normal response and a normal heterogeneity distribution.

Usage

```
BANOVA.Normal(l1_formula = "NA", l2_formula = "NA", data,
  id, burnin = 500, sample = 1000, thin = 1, jags = findjags())
## S3 method for class BANOVA.Normal
summary(object, ...)
## S3 method for class BANOVA.Normal
predict(object, newdata = NULL, ...)
## S3 method for class BANOVA.Normal
print(x, ...)
```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data.frame in long format including all features in level 1 and level 2(covariates and categorical factors) and responses
id	subject ID of each response unit
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.Normal (returned by BANOVA.Normal)
newdata	test data, either a matrix, vector or a data frame. It must have the same format with the original data (the same column number)
x	object of class BANOVA.Normal (returned by BANOVA.Normal)
...	additional arguments,currently ignored

Details

Level 1 model:

$$y_i \sim \text{Normal}(\eta_i, \sigma)$$

where $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of response i . see [BANOVA-package](#)

Value

BANOVA.Normal returns an object of class "BANOVA.Normal". The returned object is a list containing:

anova.table	table of effect sizes BANova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
data	original data.frame
mf1	model.frame of level 1
mf2	model.frame of level 2
JAGSmodel	JAGS model

Examples

```
## Not run:
# Use the ipadstudy data set
data(ipadstudy)
# mean center covariates
ipadstudy$age <- ipadstudy$age - mean(ipadstudy$age)
ipadstudy$owner <- ipadstudy$owner - mean(ipadstudy$owner)
ipadstudy$gender <- ipadstudy$gender - mean(ipadstudy$gender)
res <- BANOVA.Normal(attitude~1, ~owner + age + gender + selfbrand*conspic, ipadstudy,
ipadstudy$id, burnin = 5000, sample = 1000, thin = 10 )

summary(res)

## End(Not run)
```

BANOVA.ordMultinomial *Estimation of BANOVA with a ordered Multinomial response variable*

Description

BANOVA.ordMultinomial implements a Hierarchical Bayesian ANOVA for ordered multinomial responses, with a normal heterogeneity distribution.

Usage

```

BANOVA.ordMultinomial(l1_formula = "NA",
  l2_formula = "NA", data, id, burnin = 500, sample = 1000, thin = 1, jags = findjags())
## S3 method for class BANOVA.ordMultinomial
summary(object, ...)
## S3 method for class BANOVA.ordMultinomial
predict(object, newdata = NULL, ...)
## S3 method for class BANOVA.ordMultinomial
print(x, ...)

```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data frame
id	subject ID of each response unit
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.ordMultinomial (returned by BANOVA.ordMultinomial)
newdata	test data, either a matrix, vector or a data frame. It must have the same format with the original data (the same column number)
x	object of class BANOVA.ordMultinomial (returned by BANOVA.ordMultinomial)
...	additional arguments, currently ignored

Details

Level 1 model:

$$y_i = 1, \text{ if } l_i < 0$$

$$y_i = 2, \text{ if } 0 < l_i < c_2$$

...

$$y_i = k, \text{ if } c_{k-1} < l_i < \infty$$

$l_i = \eta_i + \epsilon_i$ where $\epsilon_i \sim \text{logistic}(0, 1)$, c_k , ($k = 2, \dots, k-1$) are cut points,

$$\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p, \text{ } s_i \text{ is the subject id of response } i. \text{ see } \text{BANOVA-package}$$

Value

BANOVA.ordMultinomial returns an object of class "BANOVA.ordMultinomial". The returned object is a list containing:

anova.table	table of effect sizes BAnova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues

```

dMatrice      design matrices at level 1 and level 2
samples_l2_param
                posterior samples of level 2 parameters
samples_cutp_param
                posterior samples of cutpoints
data          original data.frame
mf1           model.frame of level 1
mf2           model.frame of level 2
JAGSmodel     JAGS model

```

Examples

```

## Not run:

data(goalstudy)
res <- BANOVA.ordMultinomial (perceivedsim~1, ~goalprogress*varmeans, goalstudy,
goalstudy$id, burnin = 3000, sample = 1000, thin = 5)
summary(res)
BANova(res)
table.means(res)
predict(res, data[1:5,])

## End(Not run)

```

BANOVA.Poisson

Estimation of BANOVA with Poisson dependent variables

Description

BANOVA.Poisson implements a Hierarchical Bayesian ANOVA for models with a count-data response variable and normal heterogeneity distribution.

Usage

```

BANOVA.Poisson(l1_formula = "NA", l2_formula = "NA",
  data, id, burnin = 500, sample = 1000, thin = 1, jags = findjags())
## S3 method for class BANOVA.Poisson
summary(object, ...)
## S3 method for class BANOVA.Poisson
predict(object, newdata = NULL, ...)
## S3 method for class BANOVA.Poisson
print(x, ...)

```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data.frame in long format including all features in level 1 and level 2(covariates and categorical factors) and responses
id	subject ID of each response unit
burnin	the number of burn in draws in the MCMC algorithm
sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.Poisson (returned by BANOVA.Poisson)
newdata	test data, either a matrix, vector or a data frame. It must have the same format with the original data (the same column number)
x	object of class BANOVA.Poisson (returned by BANOVA.Poisson)
...	additional arguments,currently ignored

Details

Level 1 model:

$$y_i \sim \text{Poisson}(\lambda_i), \lambda_i = \exp(\eta_i + \epsilon_i)$$

where $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of response i , see [BANOVA-package](#). ϵ_i is a dispersion term.

Value

BANOVA.Poisson returns an object of class "BANOVA.Poisson". The returned object is a list containing:

anova.table	table of effect sizes BAnova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
data	original data.frame
mf1	model.frame of level 1
mf2	model.frame of level 2
JAGSmodel	JAGS model

Examples

```
## Not run:

# use the bpnodata dataset
data(bpnodata)
# within-subjects model using the dependent variable : PIC_FIX
res1 <- BANOVA.Poisson(PIC_FIX ~ AD_ID + PIC_SIZE+ PAGE_NUM
+ PAGE_POS, ~1, bpnodata, bpnodata$RESPONDENT_ID)
summary(res1)

# use the goalstudy dataset
data(goalstudy)
goalstudy$bid <- as.integer(goalstudy$bid + 0.5)
res2<-BANOVA.Poisson(bid~1, ~goalprogress*varmeans, goalstudy, goalstudy$id,
burnin = 5000, sample = 1000, thin = 20)
summary(res2)

## End(Not run)
```

BANOVA.T

*Estimation of BANOVA with T-distributin of the dependent variable***Description**

BANOVA.T implements a Hierarchical Bayesian ANOVA for linear models with T-distributed response.

Usage

```
BANOVA.T(l1_formula = "NA", l2_formula = "NA", data, id,
  burnin = 1000, sample = 1000, thin = 2, jags = findjags())
## S3 method for class BANOVA.T
summary(object, ...)
## S3 method for class BANOVA.T
predict(object, newdata = NULL,...)
## S3 method for class BANOVA.T
print(x, ...)
```

Arguments

l1_formula	formula for level 1 e.g. 'Y~X1+X2'
l2_formula	formula for level 2 e.g. '~Z1+Z2', response variable must not be included
data	a data.frame in long format including all features in level 1 and level 2(covariates and categorical factors) and responses
id	subject ID of each response unit
burnin	the number of burn in draws in the MCMC algorithm

sample	target samples in the MCMC algorithm after thinning
thin	the number of samples in the MCMC algorithm that needs to be thinned
jags	the system call or path for activating JAGS. Default calls findjags() to attempt to locate JAGS on your system
object	object of class BANOVA.T (returned by BANOVA.T)
newdata	test data, either a matrix, vector or a data frame. It must have the same format with the original data (the same column number)
x	object of class BANOVA.T (returned by BANOVA.T)
...	additional arguments, currently ignored

Details

Level 1 model:

$$y_i \sim t(\nu, \eta_i, \sigma^{-2})$$

where $\eta_i = \sum_{p=0}^P \sum_{j=1}^{J_p} X_{i,j}^p \beta_{j,s_i}^p$, s_i is the subject id of response i , see [BANOVA-package](#). The hyper parameters: ν is the degree of freedom, $\nu \sim \text{Piosson}(1)$ and σ is the scale parameter, $\sigma \sim \text{Gamma}(1,1)$.

Value

BANOVA.T returns an object of class "BANOVA.T". The returned object is a list containing:

anova.table	table of effect sizes BAnova
coef.tables	table of estimated coefficients
pvalue.table	table of p-values table.pvalues
dMatrice	design matrices at level 1 and level 2
samples_l2_param	posterior samples of level 2 parameters
data	original data.frame
mf1	model.frame of level 1
mf2	model.frame of level 2
JAGSmodel	JAGS model

Examples

```
## Not run:

# Use the ipadstudy data set
data(ipadstudy)
res <- BANOVA.T(attitude~1, ~owner + age + gender + selfbrand*conspic, ipadstudy,
ipadstudy$id, burnin = 3000, sample = 1000, thin = 5 )
res
summary(res)

## End(Not run)
```

bernlogtime	<i>Data for analysis of effects of typicality, blur and color on gist perception of ads</i>
-------------	---

Description

Data from a mixed design experiment, where respondents were exposed to 32 ads, for 100 millisecond. The ads were either typical or atypical (typical: 1 or 2). Respondents were exposed to ads that were either in full color or black-and-white (color: 1 or 2), and at different levels of blur (1=normal, 5 = very high blur). These are between-subjects factors. The dependent variables are the response 0/1, and the response time. Typicality is a within-subjects variable.

Usage

```
data(bernlogtime)
```

Format

This R object contains within-subject variable: `\$typical` is a factor with 2 levels "0" (typical ads) and "1" (atypical ads); between-subjects variables: `\$blur` is a factor with two levels (1=normal, 5 = very high blur). `\$color` denotes a factor with 2 levels "1" (full color) and "2" (grayscale). `\$subject` is the ID of subjects. `\$response` denotes if the ad is correctly identified. `\$logtime` is the response time.

```
\$bernlogtime: 'data.frame': 3072 obs. of 6 variables:
```

```
... \$ subject : int 5 5 5 5 5 5 5 5 5 ...
```

```
... \$ typical : Factor w/ 2 levels "1","2": 1 2 1 1 1 2 2 2 1 ...
```

```
... \$ blur : Factor w/ 2 levels "1","5": 1 1 1 1 1 1 1 1 1 ...
```

```
... \$ color : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 ...
```

```
... \$ response: int 1 1 1 1 1 1 1 1 1 ...
```

```
... \$ logtime : num 0.977 1.73 1.784 1 1.149 ...
```

References

Wedel, M. and Pieters, R. *The Role of Color Schemes when Advertising Exposures are Brief and Blurred*. Working paper, Robert H. Smith School of Business, University of Maryland, May 2014.

Examples

```
## Not run:
data(bernlogtime)

# model using the dependent variable : log of the response time(logtime)
res1 <- BANOVA.Normal(logtime~typical, ~blur + color, bernlogtime,
bernlogtime$subject)
res1
summary(res1)
table.means(res1)
```

```
# model using the dependent variable : response
res2 <- BANOVA.Bern(response~typical, ~blur + color, bernlogtime,
bernlogtime$subject)
summary(res2)
table.means(res2)

## End(Not run)
```

bpndata

Eye-movement data for analysis of print ad designs

Description

Data were collected in an experimental study in which 88 participants freely paged through a magazine at home or in a waiting room. While flipping through pages at their own pace, participants' eye-movements were recorded with infra-red corneal reflection eye-tracking methodology. In a subsequent memory task, participants were asked to identify the target brand in the ad as soon as possible by touching the correct brand name on the screen. **Accuracy** (accurate=1, inaccurate =0) of brand memory and **response time** were recorded for each ad and participant.

Usage

```
data(bpndata)
```

Format

This R object contains 3080 observations in the data (35 ads x 88 participants). The goal is to examine the effects of several ad design variables on both eye movements and memory. The variables include:

1. RESPONDENT_ID: ID number of a respondent;
2. AD_ID: ID number of an ad;
3. PAGE_NUM: page number in the magazine where an ad appears (1,2,3,...);
4. PAGE_POS: the right-side vs. left-side position on a page, 1 = right, 0 = left;
5. PIC_FIX: fixation count of the pictorial element (0, 1, 2, 3, ...);
6. PIC_SIZE: surface size of the pictorial element, in inches²;
7. RECALL_ACCU: whether a respondent accurately recalls the brand name, 1= yes, 0 = no;
8. RECALL_TIME: the time it takes a respondent to answer the brand recall question, in seconds.

```
\$ bpndata: 'data.frame': 3080 obs. of 8 variables:
... \$ RESPONDENT_ID: int 1 1 1 1 1 1 1 1 1 1 ...
... \$ AD_ID : int 1 2 3 4 5 6 7 8 9 10 ...
... \$ PAGE_NUM : int 2 5 6 11 13 14 17 18 21 22 ...
... \$ PAGE_POS : int 0 1 0 1 1 0 1 0 1 0 ...
... \$ PIC_FIX : int 0 2 1 1 1 2 0 3 3 8 ...
... \$ PIC_SIZE : num 74.2 52.6 77.6 71.4 52.4 ...
```

```
... \ $ RECALL_ACCU : int 0 0 0 0 0 0 1 1 0 0 ...
... \ $ RECALL_TIME : num 2.56 1.04 2.76 2.8 2.28 2.32 2.04 2.04 2.48 0.6 ...
```

References

Wedel, M. and Pieters, R. (Autumn, 2000). *Eye Fixations on Advertisements and Memory for Brands: A Model and Findings*, Marketing Science, Vol. 19, No. 4, pp. 297-312

Examples

```
## Not run:
data(bpndata)

# within-subjects model using the dependent variable : PIC_FIX
res1 <- BANOVA.Poisson(PIC_FIX ~ PIC_SIZE+ PAGE_NUM
+ PAGE_POS, ~1, bpndata, bpndata$RESPONDENT_ID)
res1

# within-subjects model using the dependent variable : RECALL_ACCU
res2 <- BANOVA.Bern(RECALL_ACCU~ RECALL_TIME + PAGE_NUM
+ PAGE_POS, ~1, bpndata, bpndata$RESPONDENT_ID)
res2

## End(Not run)
```

choicedata

Household Panel Data on Margarine Purchases

Description

Panel data on purchases of margarine by 204 households. Demographic variables are included.

Usage

```
data(choicedata)
```

Format

This is an R object that contains within-subjects variables and between-subjects variables:

```
\ $ choicePrice: 'data.frame': 1500 obs. of 13 variables:
... \ $ hhid : int 2100016 2100016 2100016 2100016
... \ $ choice : int 1 1 1 1 1 4 1 1 4 1
```

Within-subject variables:

```
... \ $ PPK\_Stk : num 0.66 0.63 0.29 0.62 0.5 0.58 0.29 ...
... \ $ PBB\_Stk : num 0.67 0.67 0.5 0.61 0.58 0.45 0.51 ...
```

```
... \ $ PFI\_Stk : num 1.09 0.99 0.99 0.99 0.99 0.99 0.99 ...
... \ $ PHse\_Stk: num 0.57 0.57 0.57 0.57 0.45 0.45 0.29 ...
... \ $ PGen\_Stk: num 0.36 0.36 0.36 0.36 0.33 0.33 0.33 ...
... \ $ PSS\_Tub : num 0.85 0.85 0.79 0.85 0.85 0.85 0.85 ...
```

Pk is Parkay; BB is BlueBonnett, Fl is Fleischmanns, Hse is house, Gen is generic, SS is Shed Spread. _Stk indicates stick, _Tub indicates Tub form.

Between-subject variables:

```
... \ $ Income : num 32.5 17.5 37.5 17.5 87.5 12.5 ...
... \ $ Fam\_Size : int 2 3 2 1 1 2 2 2 5 2 ...
... \ $ college : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 ...
... \ $ whtcollar: Factor w/ 2 levels "0","1": 0 0 0 0 0 0 1 1 1 ...
... \ $ retired : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 ...
```

Details

choice is a multinomial indicator of one of the 6 brands (in order listed under format). All prices are in \$.

Source

Allenby, G. and Rossi, P. (1991), *Quality Perceptions and Asymmetric Switching Between Brands*, *Marketing Science*, Vol. 10, No.3, pp. 185-205.

References

Chapter 5, *Bayesian Statistics and Marketing* by Rossi et al.
<http://www.perossi.org/home/bsm-1>

Examples

```
## Not run:
data(choicedata)
# generate dataX(convert the within-subjects variables to a list)
dataX <- list()
for (i in 1:nrow(choicedata)){
  logP <- as.numeric(log(choicedata[i,3:8]))
  # Note: Before the model initialization, all numeric variables(covariates) must be mean centered
  dataX[[i]] <- as.data.frame(logP) - mean(logP)
}

dataZ <- choicedata[,9:13]
res <- BANOVA.Multinomial(~ logP, ~ college, dataX, dataZ, choicedata$choice, choicedata$hhid)
summary(res)
predict(res,dataX[1:4], dataZ[1:4,])

## End(Not run)
```

 colorad

Data for gist perception of advertising, study 1

Description

Data from an experiment in which one hundred and sixteen subjects (53 men; mean age 23, ranging from 21 to 28) 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, the categories being car, financial services, food, and skincare. Subjects were asked to identify each image being flashed for 100msec. as being an ad or not. The total number of correct ad identifications, for typical and atypical ads, are used as a dependent variable.

Usage

```
data(colorad)
```

Format

This R object contains within-subject variable `\$typic` which is a factor with 2 levels "0" (typical ads) and "1"(atypical ads); between-subjects variables: `\$blur` which is a numerical variable denotes 5 different levels of blur (which must be mean centered), `\$blurfac` is a categorical data corresponding to the levels of `\$blur`, `\$color` which is a factor with 2 levels "0"(full color) and "1"(grayscale). `\$id` is the ID of subjects. `\$y` is the number of correct identifications of the 16 ads of each subject for each level of `\$typic`.

```
\$ colorad: 'data.frame': 474 obs. of 8 variables:
... \$ id : int 1 1 2 2 3 3 4 4 5 5 ...
... \$ typic : Factor w/ 2 levels "0","1": 0 1 0 1 0 1 0 1 0 1 ...
... \$ y : int 8 6 12 6 11 9 9 11 14 14 ...
... \$ blurfac : Factor w/ 5 levels "1","2","3","4",...: 2 2 4 4 2 2 3 3 1 1 ...
... \$ color : Factor w/ 2 levels "0","1": 1 1 0 0 0 0 0 0 1 1 ...
... \$ blur: num 3.69 3.69 4.79 4.79 3.69 ...
```

References

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

Examples

```
## Not run:
data(colorad)
# mean center blur
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y ~ typic, ~ color + blur + color*blur, colorad, colorad$id, as.integer(16))
res
```

```
summary(res)
table.means(res)
# use the the first row of data to predict,
# only the in-model variables data will be used, the others will be ignored
predict(res, c(1, 0, 8, 8, 2, 8, 1, 0.03400759))

## End(Not run)
```

colorad2

Data for gist perception of advertising, study 2

Description

Data from an experiment in which One hundred and forty eight subjects (71 men; age ranging from 21 to 28) were randomly assigned to one condition of a 2 (blur: normal, very high) x 2 (color: full color, grayscale, inverted) between-participants design. Participants were exposed to 25 ads for five brands in each of five categories. Ads were selected to be typical for the category, using the same procedure as in [colorad](#). The product categories used were cars, financial services, food, skincare and fragrance. Images were flashed for 100 msec. and subjects were asked to identify whether the image was an ad or not, and if they identified it correctly as an ad, they were asked to indicate which category (out of five) was advertised. The total number of correct ad identifications and category identifications are used as dependent variables.

Usage

```
data(colorad2)
```

Format

This R object contains between-subjects variables: `\$B` is a factor corresponding to the levels of blur (normal = 0, very high = 1), `\$C1` and `\$C2` are dummy variables denote 'grayscale' and 'inverted' levels of color. `\$C` is the original factor denote the color with 3 levels. `\$ID` is the ID of subjects. `\$Y1` is the number of correct identifications of the 25 ads of each subject. `\$Y2` is the number of correct identifications of the category, given the number of correct ad identifications.

```
\$ colorad2: 'data.frame': 148 obs. of 7 variables:
```

```
... \$ ID : int 1 2 3 4 5 6 7 8 9 10 ...
```

```
... \$ C1 : int 0 1 1 0 0 0 0 0 1 1 ...
```

```
... \$ C2 : int 0 0 0 1 1 0 0 0 0 0 ...
```

```
... \$ B : Factor w/ 2 levels "0","1": 1 1 0 0 1 0 0 1 0 1 ...
```

```
... \$ Y1 : int 14 6 23 21 8 23 24 5 23 6 ...
```

```
... \$ Y2 : int 2 3 8 8 2 15 10 1 13 0 ...
```

```
... \$ C : Factor w/ 3 levels "1","2","3": 1 2 2 3 3 1 1 1 2 2 ...
```

References

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

Examples

```
## Not run:
data(colorad2)
# factor C is effect coded
res1 <- BANOVA.Bin(Y1 ~ 1, ~ C + B + C*B, colorad2, colorad2$ID, as.integer(25))
# or use dummy variables
# res1 <- BANOVA.Bin(Y1 ~ 1, ~ C1 + B + C2 + C1*B + C2*B, colorad2, colorad2$ID, as.integer(25))
res1
summary(res1)
table.means(res1)

# second model
res2 <- BANOVA.Bin(Y2 ~ 1, ~ C + B + C*B, colorad2, colorad2$ID, colorad2$Y1)
summary(res2)
table.means(res2)

## End(Not run)
```

conv.diag

*Function to display the convergence diagnostics***Description**

The Geweke diagnostic and the Heidelberg and Welch diagnostic are reported. These two convergence diagnostics are calculated based on only a single MCMC chain. Both diagnostics require a single chain and may be applied with any MCMC method. The functions [geweke.diag](#), [heidel.diag](#) in **coda** package is used to compute this diagnostic.

Geweke's convergence diagnostic is calculated by taking the difference between the means from the first n_A iterations and the last n_B iterations. If the ratios n_A/n and n_B/n are fixed and $n_A + n_B < n$, then by the central limit theorem, the distribution of this diagnostic approaches a standard normal as n tends to infinity. In our package, $n_A = .2 * n$ and $n_B = .5 * n$.

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 stationary test. The convergence test uses the Cramer-von Mises statistic to test for stationary. The test is successively applied on the chain. If the null hypothesis is rejected, the first 10% of the iterations are discarded and the stationarity test repeated. If the stationary test fails again, an additional 10% of the iterations are discarded and the test repeated again. The process continues until 50% of the iterations have been discarded and the test still rejects. In our package, $eps = 0.1$, $pvalue = 0.05$ are used as parameters of the function [heidel.diag](#).

Usage

```
conv.diag(x)
```

Arguments

x the object from BANOVA.*

Value

conv.diag returns a list of two diagnostics:

```
sol_geweke      The Geweke diagnostic
sol_heidel      The Heidelberg and Welch diagnostic
```

References

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

Geweke, J. *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.

Heidelberger, P. and Welch, PD. (1981) *A spectral method for confidence interval generation and run length control in simulations*, Comm. ACM. Vol. 24, No.4, pp. 233-245.

Heidelberger, P. and Welch, PD. (1983) *Simulation run length control in the presence of an initial transient*, Opns Res., Vol.31, No.6, pp. 1109-44.

Schruben, LW. (1982) *Detecting initialization bias in simulation experiments*, Opns. Res., Vol. 30, No.3, pp. 569-590.

Examples

```
## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
summary(res)

conv.diag(res)

## End(Not run)
```

goalstudy

Data for the study of the impact of the variety among means on motivation to pursue a goal

Description

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 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, and the bid they were willing to make for the bars, used as dependent variables in the study.

Usage

```
data(goalstudy)
```

Format

This R object contains between-subjects variables: 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 set of 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 .

```
\$ goalstudy: 'data.frame': 105 obs. of 5 variables:
... \ $ id : int 1 2 3 4 5 6 7 8 9 10 ...
... \ $ perceivedsim : int 5 7 2 2 5 5 5 4 5 7 ...
... \ $ goalprogress : Factor w/ 2 levels "1","2": 1 1 2 2 2 1 2 1 2 1 ...
... \ $ varmeans : Factor w/ 2 levels "1","2": 2 1 2 1 1 1 1 2 1 1 ...
... \ $ bid : num 5 0 1 15 3 10 5 4.5 3 0.75 ...
```

References

Etkin, J. and Ratner, R. (2012) *The Dynamic Impact of Variety among Means on Motivation*. Journal of Consumer Research, Vol. 38, No. 6, pp. 1076 - 1092.

Examples

```
## Not run:
data(goalstudy)

res1 <- BANOVA.Normal(bid~1, ~goalprogress*varmeans, goalstudy, goalstudy$id,
burnin = 5000, sample = 1000, thin = 20)

conv.diag(res1)
trace.plot(res1)
summary(res1)

res2 <- BANOVA.ordMultinomial (perceivedsim~1, ~goalprogress*varmeans, goalstudy,
goalstudy$id, burnin = 3000, sample = 1000, thin = 5)

summary(res2)
table.means(res2)
# print JAGS code
cat(res2$JAGSmodel)

## End(Not run)
```

ipadstudy

Data for the study of relation between Conspicuous, Brand Usage, Self-Brand Connection and attitudes toward the brand

Description

The study is a between-subjects experiment which has factor (conspicuousness: low vs. high) and one measured variable (self-brand connection). The goal is to show that conspicuous brand use negatively affects attitudes toward the user and the brand only for observers low in self-brand connection. One hundred fifty-four participants were exposed to a video manipulating conspicuous brand usage. Participants completed the study by answering several questions which are used to measure the dependent (attitude) and independent (self-brand connection) variables in the model.

Usage

```
data(ipadstudy)
```

Format

This R object contains between-subjects variables: `\$owner` is an indicator variable. If the subject owns iPad or iPhone, then `owner = 1`. It is equal to 0 otherwise. `\$age` denotes the age of subjects. `\$gender` denotes the gender of subjects. `gender = 1` if the subject is a female, 0 otherwise. `\$conspic` is an indicator variable related to conspicuousness. `conspic = 1` if conspicuousness is high. `\$selfbrand` denotes the self-brand connection for Apple. `\$id` is the id of subjects. `\$attitude` denotes the attitudes towards the brand which is the continuous dependent variable. `\$apple_dl` is a seven-point scale variable which denotes the attitudes (dislike = 1,..., like = 7)

```
\$ ipadstudy: 'data.frame': 154 obs. of 9 variables:
... \$ id : int 1 2 3 4 5 6 7 8 9 10 ...
... \$ attitude : num 3 5.33 5.67 5.33 6 ...
... \$ owner : num 0 0 0 1 1 0 1 0 1 0 ...
... \$ age : int 19 33 25 41 38 33 37 46 41 55 ...
... \$ gender : num 0 0 1 0 1 1 1 0 1 1 ...
... \$ conspic : num 0 1 0 1 1 0 0 1 0 1 ...
... \$ selfbrand : num -2.304 1.696 -0.161 -0.447 0.267 ...
... \$ apple_dl : int 3 6 6 5 6 4 7 7 5 5 ...
```

References

Ferraro,R., Kirmani, A. and Matherly, T., (2013) *Look at Me! Look at Me! Conspicuous Brand Usage, Self-Brand Connection, and Dilution*. Journal of Marketing Research, Vol. 50, No. 4, pp. 477-488.

Examples

```
## Not run:
data(ipadstudy)
```

```

# mean center covariates
ipadstudy$age <- ipadstudy$age - mean(ipadstudy$age)
ipadstudy$owner <- ipadstudy$owner - mean(ipadstudy$owner )
ipadstudy$gender <- ipadstudy$gender - mean(ipadstudy$gender)

res <- BANOVA.Normal(attitude~1, ~owner + age + gender + selfbrand*conspic,
ipadstudy, ipadstudy$id, burnin = 3000, sample = 1000, thin = 5 )
summary(res)

# use apple_dl as the dependent variable
res <- BANOVA.ordMultinomial(apple_dl~1, ~owner + age + gender + selfbrand*conspic,
ipadstudy, ipadstudy$id, burnin = 3000, sample = 1000, thin = 5 )
summary(res)
table.means(res)

## End(Not run)

```

table.means

Function to print the table of means

Description

Output of this function is a table of means for the categorical predictors at either level-1 or level-2. Statistics of interest such as credible intervals and standard deviations of the means are also computed.

Usage

```
table.means(x)
```

Arguments

x the object from BANOVA.*

Examples

```

## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
summary(res)
table.means(res)

## End(Not run)

```

table.pvalues

*Function to print the table of p-values***Description**

Computes the Bayesian p-values for the test concerning the level 2 parameters:

For $p = 1, \dots, P$

$$H_0 : \theta_{j,k}^{p,q} = 0$$

$$H_1 : \theta_{j,k}^{p,q} \neq 0$$

The two-sided P-value for the sample outcome is obtained by first finding the one sided P-value, $\min(P(\theta_{j,k}^{p,q} < 0), P(\theta_{j,k}^{p,q} > 0))$ which can be estimated from posterior samples. For example, $P(\theta_{j,k}^{p,q} > 0) = \frac{n_+}{n}$, where n_+ is the number of posterior samples that are greater than 0, n is the target sample size. The two sided P-value is $P_\theta(\theta_{j,k}^{p,q}) = 2 * \min(P(\theta_{j,k}^{p,q} < 0), P(\theta_{j,k}^{p,q} > 0))$.

If there are $\theta_{j,k_1}^{p,q}, \theta_{j,k_2}^{p,q}, \dots, \theta_{j,k_J}^{p,q}$ representing J levels of a multi-level variable, we use a single P-value to represent the significance of all levels. The two alternatives are:

$$H_0 : \theta_{j,k_1}^{p,q} = \theta_{j,k_2}^{p,q} = \dots = \theta_{j,k_J}^{p,q} = 0$$

$$H_1 : \text{some } \theta_{j,k_j}^{p,q} \neq 0$$

Let $\theta_{j,k_{min}}^{p,q}$ and $\theta_{j,k_{max}}^{p,q}$ denote the coefficients with the smallest and largest posterior mean. Then the overall P-value is defined as

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

```
table.pvalues(x)
```

Arguments

x the object from BANOVA.*

Source

It borrows the idea of Sheffe F-test for multiple testing: the F-stat for testing the contrast with maximal difference from zero. Thank Dr. P. Lenk of the University of Michigan for this suggestion.

Examples

```
## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
summary(res)
```

```
table.pvalues(res)

## End(Not run)
```

trace.plot	<i>Function to plot the trace of parameters</i>
------------	---

Description

Function to plot the trace of all level-2 parameters. The plots can be saved as a pdf file.

Usage

```
trace.plot(x, save = FALSE)
```

Arguments

x	the object from BANOVA.*
save	whether to save the trace plot as a pdf file, the default is FALSE

Examples

```
## Not run:
data(colorad)
# mean center Blur for effect coding
colorad$blur <- colorad$blur - mean(colorad$blur)
res <- BANOVA.Bin(y~typic, ~color*blur, colorad, colorad$id, as.integer(16),
burnin = 3000, sample = 2000, thin = 5)
trace.plot(res)

## End(Not run)
```

Chapter C: **BANOVA** Web Application

In this appendix, we introduce a web application of Bayesian ANOVA (address: <http://chendong.physics.umd.edu/>). The motivation for the development of the web application is to offer user-friendly menu driven analysis of Bayesian hierarchical ANOVA models introduced in Chapter 2.

To use the web application, user needs to first log in using a guest account provided on the web page. Once logged in, user can start building a Bayesian ANOVA model by uploading a data set first (see Figure C.1). The data set must be in a CSV(comma-separated values) file which follows the long format introduced in Figure 2.1. As introduced before, to successfully build a Bayesian ANOVA model, the data set must include a column indicating the subject ID, a column indicating the response variable, columns indicating within-subject and between-subject variables.

The web application automatically reads the data and then let users select several in-model variables. Using the `colorad` data introduced in Table 2.3 as an example, the whole process includes 6 steps:

1. Select the subject ID variable (Figure C.2). The subject ID variable is very important to the Bayesian ANOVA model, since it contains the information related to the heterogeneity within and between different subjects;
2. Select a dependent variable and the corresponding Bayesian ANOVA model

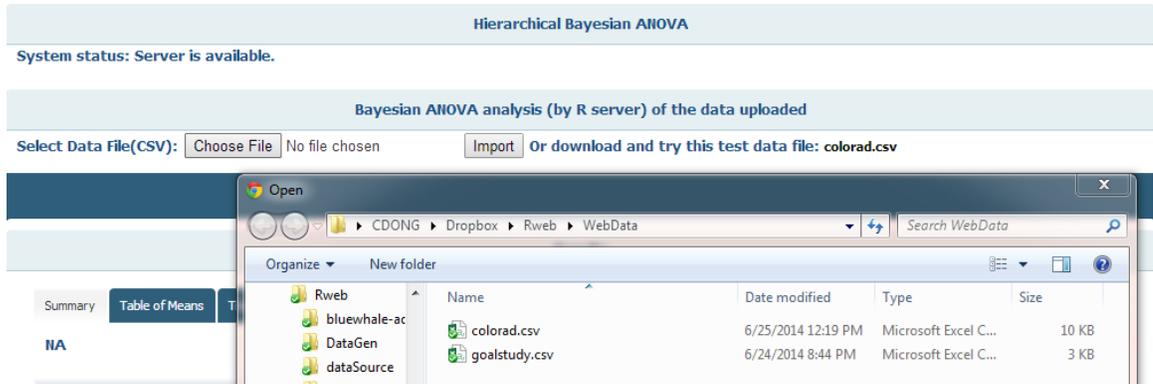


Figure C.1: Data input

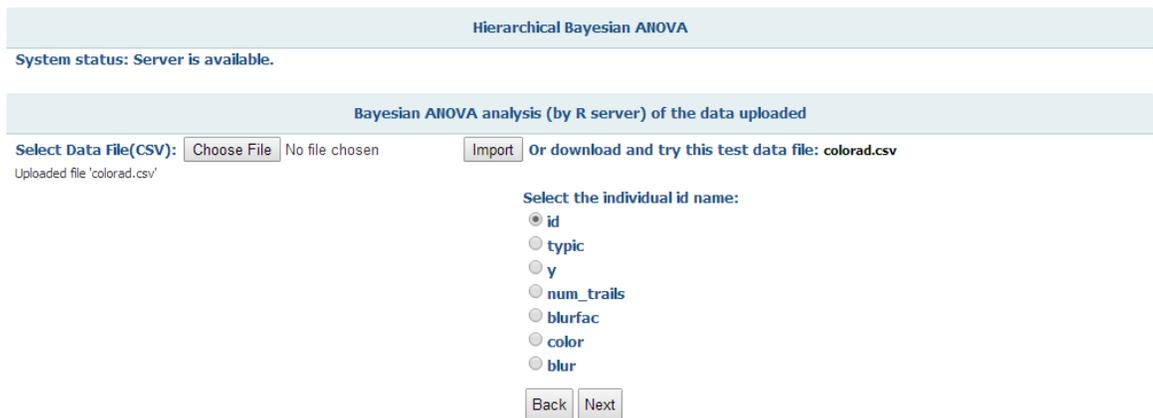


Figure C.2: Select subject ID

(Figure C.3). Currently, the web application contains 6 Bayesian ANOVA models including (Normal, Bernoulli, Binomial, t, Poisson and ordered multinomial models). The assumption of the distribution of the response variable must be checked by the user before the selection. For example, if user selects the Bernoulli model but the response variable actually follows a Normal distribution (continuous). The program will report an error indicating this problem after starting the simulation.

3. Select between-subject variables and their interactions (Figure C.4 and

Hierarchical Bayesian ANOVA

System status: Server is available.

Bayesian ANOVA analysis (by R server) of the data uploaded

Select Data File(CSV): No file chosen Or download and try this test data file: [colorad.csv](#)

Select a Dependent Variable and its data type:

DV names:

- typic
- y
- num_trails
- blurfac
- color
- blur

DV data type:

- Normal distribution
- t distribution(Robust,better to handle outliers)
- Bernoulli distribution (i.e. 0,1)
- Poisson distribution (i.e. 0,1,2,3,...)
- Ordered Multinomial (i.e. 1,2,3,...)
- Binomial distribution (need total number of trails)

Figure C.3: Select the dependent variable and a model

Figure C.5). If there is no between-subject variable, users can select 'No between subjects data'. Variables that only contain few levels are considered as categorical variables which should be specified by the user. The program will deal with these variables using effect coding but not dummy coding. The choices of interactions between categorical variables will be displayed next and let users to chose. The interactions are currently limited to 3-way by the program. Note that, if users select the Binomial model, the variable indicating the total number of trails must be selected at this step.

4. Similar to the selections of between-subject variables, within-subject variables and their interactions (Figure C.6 and top part of Figure C.7) are selected at this step.

5. After selections of all in-model variables, parameters for MCMC simulation

Hierarchical Bayesian ANOVA

System status: Server is available.

Bayesian ANOVA analysis (by R server) of the data uploaded

Select Data File(CSV): No file chosen Or download and try this test data file: [colorad.csv](#)

Select total number of trails for the Binomial model:

Select between subjects data names:

Continuous

Continuous

Categorical

Categorical

Continuous

No between subjects data

Figure C.4: Select between-subject variables

Hierarchical Bayesian ANOVA

System status: Server is available.

Bayesian ANOVA analysis (by R server) of the data uploaded

Select Data File(CSV): No file chosen Or download and try this test data file: [colorad.csv](#)

Select interactions(limited to at most 3-way interactions):

blurfac*color

Figure C.5: Select interactions among between-subject variables

must be selected before the simulation(Figure C.7). The parameters include the burn-in steps (default 5000), the number of samples chosen (default 1000) and the thin factor (default 10). When all parameters are set, the program will automatically run by clicking the 'Analyze' button. An error report will be generated if there is any problem happening during the simulation. Otherwise, a report of the result will be displayed below.

6. Access the simulation results (Figure C.8). The results include three parts

Hierarchical Bayesian ANOVA

System status: Server is available.

Bayesian ANOVA analysis (by R server) of the data uploaded

Select Data File(CSV): No file chosen Or download and try this test data file: [colorad.csv](#)

Select within subjects types and names:

- Categorical
- Continuous
- Continuous
- No within subjects data

Figure C.6: Select within-subject variables

displayed by three tabs: a summary, a table of means and trace plots.

Hierarchical Bayesian ANOVA

System status: Server is available.

Bayesian ANOVA analysis (by R server) of the data uploaded

Select Data File(CSV): No file chosen Or download and try this test data file: [colorad.csv](#)

Select interactions(limited to at most 3-way interactions):
No interactions (number of categorical variable < 2).

Input simulation parameters:

Burn in iterations:(default: 5000)

Sample iterations:(default: 1000)

Thin:(default: 10)

Figure C.7: Parameters setting

Results

Summary **Table of Means** Trace Plots

Call:
BANOVA.Bin(l1_formula = y ~ 1 + typic, l2_formula = ~1 + blurfac + color + blurfac * color, data = data_long, id = data_long[, ind], num_trails = num_trails, burnin = 1000, sample = 1000, thin = 5)

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

	blurfac	color	blurfac:color	Residuals	Total
(Intercept)	30.932 (18.34%)	5.714 (3.39%)	3.264 (1.94%)	128.77	168.649
typic1	13.402 (69.09%)	2.322 (11.97%)	2.425 (12.5%)	1.354	19.399

Table of p-values (Multidimensional):

	(Intercept)	blurfac	color	blurfac:color
(Intercept)	0	0	0.008	0.152
typic	0	0	0.000	0.010

Table of coefficients:

	mean	SD	Quantile0.025	Quantile0.975
(Intercept) : (Intercept)	0.49026	0.05701	0.37573	0.59438

Figure C.8: Results display

Bibliography

- Baron RM, Kenny DA (1986). “The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations.” *Journal of Personality and Social Psychology*, **51**(6), 1173–1182.
- Bartlett M (1947). “The use of Transformations.” *Biometrics*, **3**, 39–52.
- Bauer DJ, Preacher KJ, Gil KM (2006). “Conceptualizing and Testing Random Indirect Effects and Moderated Mediation in Multilevel Models: New Procedures and Recommendations.” *Psychological Methods*, **11**, 142–163.
- Bernardo J, Smith A (2000). *Bayesian Theory*. John Wiley and Sons, West Sussex, England.
- 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.
- 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*).
- Dong C, Wedel M (2014a). “Bayesian mediation models.” *Working paper, University of Maryland*.
- Dong C, Wedel M (2014b). “**BANOVA**: An R Package for Hierarchical Bayesian ANOVA.” *Working paper, University of Maryland*.
- 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.
- Gelman A, Hill J (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY.
- 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.
- Hayes AF, Preacher KJ (2011). “Indirect and Direct Effects of a Multicategorical Causal Agent in Statistical Mediation Analysis.” *Unpublished manuscript*.
- Heidelberger P, Welch P (1983). “Simulation Run Length Control in the Presence of an Initial Transient.” *Opns Res.*, **31**, 1109–1144.
- Iacobucci D (2001). “Special Issue on the Methodological Questions of the Experimental Behavioral Researcher.” *Journal of Consumer Psychology*, **10**(1/2), 83–100.
- Iacobucci D (2012). “Mediation Analysis and Categorical Variables: The Final Frontier.” *Journal of Consumer Psychology*, **22**(4), 582–594.
- Ito P (1980). *Robustness of ANOVA and MANOVA test procedures*. In Handbook of Statistics, Vol. 1, (P. R. Krishnainh, ed.), Amsterdam, Holand.
- James LR, Brett J (1984). “Mediators, Moderators, and Tests for Mediation.” *Journal of Applied Psychology*, **69**(2), 307–321.

- Kenny DA, Kashy DA, Bolger N (1998). *Data Analysis in Social Psychology*. In D. Gilbert, S. T. Fiske, and G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 223–265). McGraw-Hill, New York, NY.
- Kenny DA, Korchmaros JD, Bolger N (2003). “Lower Level Mediation in Multilevel Models.” *Psychological Methods*, **8**, 115–128.
- Kirk RE (1982). *Experimental design: Procedures for the Behavioral Sciences*. 2 edition. Brooks/Cole, Belmont, CA.
- Kraemer HC, Wilson GT, Fairburn CG, Agras WS (2002). “Mediators and Moderators of Treatment Effects in RCTs.” *Archives of General Psychiatry*, **59**, 877–883.
- Krull JL, MacKinnon DP (1999). “Multilevel Mediation Modeling in Group-based Intervention Studies.” *Evaluation Review*, **23**, 418–444.
- Krull JL, MacKinnon DP (2001). “Multilevel Modeling of Individual and Group Level Mediated Effects.” *Multivariate Behavioral Research*, **36**, 249–277.
- 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 Stimulation 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.
- MacKinnon DP (2008). *Introduction to Statistical Mediation Analysis*. Erlbaum, New York, NY.
- MacKinnon DP, Dwyer JH (1993). “Estimating Mediating Effects in Prevention Studies.” *Evaluation Review*, **17**, 144–158.
- MacKinnon DP, Fairchild AJ, Fritz MS (2007). “Mediation Analysis.” *Annual Review of Psychology*, **58**, 593–614.
- MacKinnon DP, Krull JL, Lockwood CM (2000). “Equivalence of the Mediation, Confounding and Suppression Effects.” *Prevention Science*, **1**, 173–181.
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V (2002). “A Comparison of Methods to Test Mediation and other Intervening Variable Effects.” *Psychological Methods*, **7**(1), 83–104.
- MacKinnon DP, Warsi G, Dwyer JH (1995). “A Simulation Study of Mediated Effect Measures.” *Multivariate Behavioral Research*, **30**, 41–62.
- 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.
- Morse GA, Calsyn RJ, Allen G, Kenny DA (1994). “Helping Homeless Mentally Ill People: What Variables Mediate and Moderate Program Effects?” *American Journal of Community Psychology*, **22**, 661–683.
- Nunnally 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). “Educational 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.
- Raudenbush SW, Sampson R (1999). “Assessing Direct and Indirect Effects in Multilevel Designs with Latent Variables.” *Sociological Methods and Research*, **28**, 123–153.
- Robert C, Casella G (2004). *Monte Carlo Statistical Methods*. Springer Verlag.
- 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>.
- Rossi PE, Allenby GM, McCulloch R (2006). *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics.
- 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.
- Shrout PE, Bolger N (2002). “Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations. Psychological Methods.” *Psychological Methods*, **7**, 422–445.

- Sobel ME (1982). *Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models*. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290-312). American Sociological Association, Washington, DC.
- 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*,.
- Yuan Y, MacKinnon D (2009). "Bayesian Mediation Analysis." *Psychological Methods*, **14**(4), 301–322.
- Zhang J, Wedel M, Pieters R (2009). "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis." *Journal of Marketing Research*, **46**, 669–681.

Zhao X, Lynch Jr JG, Chen Q (2010). “Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis.” *Journal of Consumer Research*, **37**(2), 197–206.