

Alternative to Null Hypothesis Significance Testing (NHST): Understanding the Bayesian Approach*

Won-Woo Park**

Hyerim Kim***

Heekyung Kim****

Gahee Shin*****

Cyril Um*****

.....

Null Hypothesis Significance Testing is the most widely used method of analysis in social science. However, there are inherent limitations of NHST. We discuss the problems of using NHST and suggest alternative ways to solve the problems, focusing on Bayesian statistics. To supplement the limitation of NHST it has been argued that providing information regarding the effect size and confidence interval are useful. Furthermore, Bayesian statistics, which is an independent stream of study, has gained attention. We introduce the basic concept of Bayesian statistics and the application of the method in organizational research.

.....

I. Introduction

Null hypothesis significance testing is the most prevalently and widely used method of

*This study was supported by the Institute of Management Research at Seoul National University.

**Professor, College of Business Administration, Seoul National University

***Seoul National University Graduate school of public administration

****Seoul National University Business School

*****Seoul National University Technology Management Economics and Policy Program

*****Seoul National University Business School

analysis in social science. With the prominence of the frequentists, NHST had been taken for granted and had been embraced without skepticism for many years. However, recently the criticism of the method has been intensified. Surprisingly enough, the controversy regarding the significance of NHST dates back to the 19th century. Numerous researchers criticized the wrong usage of NHST and its severe deficiencies in processing research (e.g., Schmidt, 1996; Nickerson, 2000). Furthermore, recognizing the lower quality of research using p-value, the Basic and Applied Social Psychology (BASP) announced that the journal would no longer publish papers analyzed only with p-values. It seems that the glory days of the NHST has finally come to an end, at least to a small portion of the scientific community.

Unfortunately, the portion of the enlightened remains trivial. The purpose of this paper is to critically review the vulnerability of the NHST and suggest alternative method to the field of organizational science. By doing so, the authors hope that the paper can create a ripple that turns into a big systematic wave, changing the perception of scientific researchers.

This paper is organized as follows: first, we review the traditional procedure of the NHST; pin point various ways the NHST might bare problems; suggest effect size analysis as complement to NHST; propose Bayesian statistics as the alternative to the NHST. Finally, we elaborate on how the Bayesian approach has been applied and interpreted in field of management. We conclude our proposal with suggesting how Bayesian analysis could be further adapted to the field of organizational science.

II. Null hypothesis significance testing and the criticisms

1. Null hypothesis significance testing (NHST)

In most scientific researches, NHST has been used to make statistical inference. Most commonly, a data set gathered by random sampling is compared to an idealized model. Null hypothesis, as its name implies, represents the idealized model that proposes no relationship between the sampled data and synthetic data. On the other hand, alternative hypothesis is a proposition of the statistical relationship that the researchers want to prove to be true.

Thus, ironically researchers want their idealized model to be rejected. For example, suppose that a researcher tries to show that an employee who gets feedback regularly from her boss outperforms an employee who does not get any feedback from her boss. In this case, a null hypothesis can be set as the average performance of employees with feedback is the same as that of employees without feedback (Naturally, the alternative hypothesis proposes the opposite).

The modern NHST is a synthesis of the test of significance suggested by R. A. Fisher and the hypothesis test introduced by J. Neyman and E. S. Pearson (Gill, 1999). By calculating p-value from test statistic, a researcher can conclude whether he can reject the null hypothesis or not. The p-value is defined as the probability of obtaining a value equal to what was actually gathered, given that the null hypothesis is true (Levine, Weber, Hullett, Park, and Lindsey, 2008). In other words the p-value measures how extreme the observation is under the case where the null hypothesis is correct. If p-value is smaller than the significance level(α) decided by the researcher, the difference is said to be statistically significant (Figure 1). The logic behind NHST is that given the null hypothesis, if the probability of getting the sample is extremely low, then we can conclude that the null hypothesis cannot be true. However, the logic of NHST often misleads people into thinking that rejecting the null is the same as accepting the alternative. The matter concerning this misperception will be discussed in the following section.

The usual process of NHST is as follows.

- 1) State the null and alternative hypothesis.

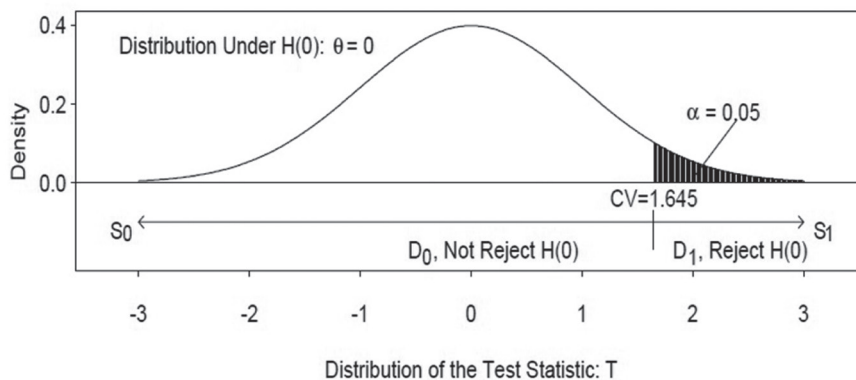


Figure 1. Null Hypothesis Significance Testing (Gill, 1999)

- 2) Determine the appropriate test statistic or the distribution under the assumption that the null hypothesis is true.
- 3) Select a significance level (α).
- 4) Calculate the test statistic from the data and then the p-value.
- 5) Reject the null hypothesis if the calculated p-value is less than the significance level chosen.

Figure 1 shows a student's t-distribution and its critical value for 5% significance level of a one-tailed test. If the computed value of test statistic is more than 1.645, then the p-value will be less than 0.05 and it is deemed statistically significant. Researchers decide which test statistic will be used in order to calculate the p-value, depending on the data they have collected. If the sample is large enough to satisfy the assumption of normal distribution, a researcher can use z-statistic. If the sample is small, then one can use the t-statistic. There are various distributions to use for different occasion and situations.

Two correct decisions are possible as the outcome of NHST reject the false and accept the truth. Accordingly, two false decisions are possible as well commonly known as type 1 error and type 2 error. Type 1 error occurs when the true null hypothesis is rejected. P-value and the significance level are related to type 1 error. Type 2 error is an error that fails to reject the false null hypothesis. Statistical power pertaining to type 2 error is the probability of rejecting the null hypothesis conditional on its being false, meaning the probability to detect an effect given that there is one (Nickerson, 2000).

Researchers are sensitive to making type 1 errors. Scientific research starts from the assumption that the null hypothesis is correct. Therefore it is crucial that they reduce the probability of rejecting the true hypothesis. In other words, by setting the significance level low, researchers can effectively control the rate of type 1 error. Conventionally, α is set at 5% or 1%. Setting the significance level to this level indicates that the probability of committing the type 1 error is set to 5% or 1% maximum. However, this conventional selection is made arbitrarily, without firm ground, causing numerous criticism to the NHST.

2. Problems with the null hypothesis significance testing

The Basic and Applied Social Psychology (BASP) 2014 Editorial emphasized that the NHST procedure is invalid (Trafimow, 2014). After allowing a grace period of using NHST, the BASP 2015 Editorial announced that the journal no longer allows the usage of NHST. It contended that the ban of NHST procedure does not mean easier publication in BASP. Rather, it mentioned that the $p < .05$ bar was too easy to pass and served as an excuse for lower quality research. The journal's intention is to increase the quality of submitted manuscripts by liberating researchers from the stultified logic of NHST, thereby, eliminating an important obstacle to creative thinking (Trafimow and Marks, 2015). In this perspective, there is a need to check what the so-called stultified structure of NHST is.

The most pervasive criticism of NHST is that NHST is not well-understood by its users and consequently they draw conclusions on the basis of test results that the data do not justify (Nickerson, 2000). The misunderstandings of the NHST are listed as follow: a belief that a small p-value means a treatment effect of large magnitude; a belief that statistical significance means theoretical or practical significance; a belief that failing to reject the null hypothesis is equivalent to demonstrating the alternative to be true; and a belief that a failure to reject the null hypothesis is evidence of a failed experiment (Nickerson, 2000).

In addition to the misconceptions mentioned above NHST is criticized because it provides relatively little information about the relationship between the dependent and independent variables (Nickerson, 2000). It only tells whether the statistically significant effect is obtained without giving a measure of the size of the effect or the strength of the relationship between variables of interest. On the contrary, as the paper will later show, Bayesian analysis provides a posterior probability for each of a set of hypotheses of interest (Cronbach and Snow, 1977).

Another drawback regarding NHST is its arbitrariness of alpha criterion (Nickerson, 2000). The most widely recommended alpha criterion is .05 and Nickerson (2000) mentioned that .05 is treated by many researchers as an upper bound on what should be considered statistically significant, but relatively few specify it as an alpha in advance of collecting data and then report all results relative to that criterion. Regarding the decision criterion, American Psychological

Association manual (1994) allows considerable latitude in the selection of an alpha level and the reporting of p-values, while others believe that such latitude adds undesirable elements of subjectivity to the process of hypothesis evaluation (Frick, 1996).

The failure to distinguish between informative and non-informative hypothesis tests is also problematic. According to Bonett and Wright (2007), the non-informative hypothesis tests do not provide new information on the unknown parameter values of the population. To illustrate, the ANOVA F-test is a non-informative hypothesis test because we already know that the null is false. Savage (1957) states that null hypotheses of no difference are usually known to be false before the data are collected, and when they are, their rejection or acceptance simply reflects the size of the sample and the power of the test, and is not a contribution to science. Moreover, in reality, the null hypothesis can never be exactly true and will therefore always be rejected as the number of observations increases (Wagenmakers, 2007).

Another problem with NHST arises from the fact that researchers want to obtain a sufficiently small p-value of $P(X \geq y | H_0)$ so as to reject the null hypothesis (where H_0 denotes the null hypothesis, X the random variable and y the observed data). Fisher (1959) knew that such a test of significance does not authorize researchers to make any statement about the hypothesis in question in terms of mathematical probability, since such statements represent $P(H_0 | X \geq y)$, which has significantly different meaning. Although Fisher (1959) stated that $P(X \geq y | H_0)$ may influence the null's acceptability, p-values without priors — unconditional probabilities — can be highly misleading measures of the evidence provided by the data against the null hypothesis (Nickerson, 2000).

Finally, from the statistical perspective, p-values depend on data that were never observed. The p-value is not only based on the test statistic for the observed data, $t(x) | H_0$, but also on hypothetical data that were never observed. With these hypothetical data expected under H_0 , we can construct the sampling distribution of the test statistic $t(X_{\text{rep}} | H_0)$. The concern over dependence on hypothetical data is not minor since it is a violation of conditionality principle which states that statistical conclusions should be based on data that have actually been observed (Wagenmakers, 2007).

Table 1. Categorization of the effect size indicants (Thompson, 2002)

	effect size related to standardized differences	effect size related to variance-accounted for
uncorrected	Glass's Δ Cohen's d	η^2
corrected	Thompson's "corrected" d	Hay's w^2 Adjusted R^2

III. Ways to supplement the NHST

1. Effect size

Nickerson (2000) suggested various ways to supplement the inadequacy of NHST, one of those being reporting the effect size. An effect size refers to a sample-based estimate of the size of the relationship between variables (Rosenthal, 1994). Cohen (1977) defined it as 'the degree to which the phenomenon is present in the population', or 'the degree to which the null hypothesis is false'. In simple terms, it is the quantitative measure of the strength of the phenomenon.

There are a variety of indicants of the effect size. As for the difference between means, there are Cohen's d , Glass's Δ , and Hedges' g ; as to the measures of association, there are correlational indicators such as Pearson's r , Spearman's ρ , and Phi coefficient, and explained variance indicators such as η^2 , ϵ^2 , w^2 , and intra-class correlation coefficient (Fern and Monroe, 1996). Among them, the most widely used indicators are r^2 , which is the coefficient of determination, and Cohen's d , which is the difference between means divided by the pooled within-groups standard deviation. Researchers (e.g., Kirk, 1996; Richardson, 1996) categorized those indicants as one of two types—measures of the standardized difference between group means and measures of explained variance.

Reporting effect size has several merits. Hagen (1998) indicated that it may be consistent with the use of NHST and an important complement to it. In addition, being calculated

regardless of the p-values, the effect size is useful in determining sample sizes for subsequent experimentation and tells something that a p-value does not (Rosnow and Rosenthal, 1989). According to Kim (2011), it is free from the NHST's problem of dichotomy, and it provides useful information for the replicability and in determining the importance of research results. Besides, it facilitates the use of meta-analytic techniques (Asher, 1993).

However, it is noteworthy that the opinions differ regarding the merits of specific possibilities of the effect size (Crow, 1991; McGraw, 1991; Rosenthal, 1991) and which indicant would be the most appropriate in a given context is not always apparent (Rosenthal, 1991; Nickerson, 2000). Accordingly, the effect size should be used properly in a given context of the research and its implications need to be interpreted carefully.

2. Confidence Interval

A confidence interval is a kind of interval estimate of the population parameter. The confidence interval is 'usually interpreted as the range of values that encompass the population or 'true' value, estimated by a certain statistic, with a given probability' (Nakagawa and Cuthill, 2007). Basically, a confidence interval can be calculated as follows.

- 1) Calculate the sample mean.
- 2) Select the confidence level and compute a test statistic. If the standard deviation is known, then calculate z-statistic. Otherwise, use t-statistic in order to get the critical value.
- 3) Compute the appropriate equation with the value obtained in the step 2.

$$CI = \begin{cases} \left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}} \right) & \text{if a standard deviation is known} \\ \left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}} \right) & \text{else} \end{cases}$$

Many critics on NHST suggest that researchers should report a confidence interval with a point estimate of an effect size (Rozeboom, 1960; Duhachek and Iacobucci, 2004; Dalton and Dalton, 2008; Edwards, 2008). Rozeboom (1960) suggested that whenever possible, the basic statistical report should be in the form of confidence interval. It is because the information

given by the confidence interval is more comprehensive than that of the solely given p-value. For example, the confidence interval can provide information about the direction as well as the magnitude of the difference in population means that is not provided with just a single hypothesis test (Bonett and Wright, 2007). We can also find out that the result is not statistically significant, if the confidence interval includes 0.

However, like NHST, confidence interval is misinterpreted by many. For example, numerous people think on the flawed notion that a 95% confidence interval indicates a 95% chance of the sample mean being in that specific interval. Unfortunately, this is wrong (Cumming, Williams, and Fidler, 2004). The true meaning of CI is not as straight forward as one might assume. Assuming that the confidence level is set at 95%, what CI is telling the researchers is that when numerous trials of calculating the CI, 95% of trial would capture the true population value. Cox and Hinkley (1979) indicated that ‘the confidence interval represents values for the population parameter for which the difference between the parameter and the observed estimate is not statistically significant’ at the set level. Researchers always have to carefully interpret and treat the result from the confidence interval.

Another criticism made by Edwards and Berry (2010) implied that the use of the confidence interval is just tantamount to conducting NHST. The confidence interval cannot resolve the paradox described by Meehl (1967, 1978). Even though there can be some misleading and unsolved problem, a confidence interval with a point estimate of an effect size can be integrated into a meta-analysis later. A meta-analysis lets researchers understand the big picture of the research subject. In meta-analysis, many different individual researches are aggregated. Thus, the confidence intervals around the meta-analytic estimates of effect size will be narrower than those of each individual study and can become more precise to know the population parameter (Carlson and Ji, 2011).

IV. Bayesian statistics as the alternative to NHST

1. Bayesian Statistics

1) Probability

The theory of statistical inference and decision making process embed the concept of uncertainty; no data or results are definitive. It is evident that when one makes decisions, he or she has to encounter uncertain outcomes of various situations, which have an important bearings to the problem. The concept of probability quantifies this uncertain aspect of outcomes and since this notion enables people to express uncertainty in quantified form, it is very crucial in inference and decision making process. In other words, probability can be thought of as the mathematical language of uncertainty (Winkler, 1972).

Then defining probability accurately is a crucial matter. Bayesian and frequentists have had debate on the meaning of probability for more than a century (Daston, 1994; Galavotti, 2005). The ongoing debate originates from an important difference in the interpretation of probability. Bayesian probability is associated with the subjective nature of uncertainty. From this approach, probability is interpreted as the degree of belief or degree of knowledge of a particular individual. On the other hand, frequentist probability is based on the features of observable systems. It is associated with the relative frequency of an event in the long run. According to the *law of large numbers*, if an experiment is repeated numerously under identical conditions, the relative frequency of an event will be the probability of that event. Formally, if n denotes the number of times the experiment was performed and r denotes the number of occurrence of event E then, we can formulate the property of event E occurring as,

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{r}{n} - P(E) \right| \geq \epsilon \right] = 0,$$

where ϵ is any arbitrarily small positive number.

In modern days, interpretation in terms of relative frequency is used more prevalently than that of its counterpart. This is because ‘classical’ interpretation was originally developed

to describe certain cases of chance where events are actually repeated for a large number of trials with equal likelihood (Winkler, 1972). Similarly, there are myriads of situations when researchers make observations under the same environment, making relative-frequency interpretation essentially reasonable to use. However, frequentists are limited to inference about the probability of events from occurring in that the long run is always unobservable and the probability calculated is disconnected from researchers' actual interest in theories and models (Eagle, 2004). Also, there are many situations that can be probabilistic but cannot be interpreted in the 'classical' way. For instance, the question 'what are the chances of raining tomorrow?' appears to be probability statement, but it is difficult to answer the question with long-run relative frequency perspective. The difficulty comes from the fact that the events of raining tomorrow are fundamentally unique. Although, information regarding the past occurrences is available, it is doubtful that the information in the form of observed frequency is representing situations exactly identical to the current conditions.

Alternatively, as said before, Bayesian probability is interpreted as a measure of degree of belief or as the quantified judgment of an individual. It is logical to assign a quantified amount to an event that involves non-repetitive situations like the chance of raining tomorrow. Thus, Bayesian can apply probability to anything that can be the subject of belief or knowledge, including hypotheses, statistical parameters and models. The Bayesian paradigm is especially powerful to the field of social sciences for mainly two reasons. First, most of the data that social scientists acquire are impossible to gather repeatedly in unlimited sample space. For example, if the data is time-series cross-sectional, we cannot think of the population expanding infinitely. Only limited cases of experiments or surveys are adequate for applying the interpretation of the frequentists. By incorporating Bayesian, we can effectively overcome the limitation of frequentist probability. Second, as we have seen in the previous section, NHST is limited in that the quantity of interest cannot be expressed in probabilistic form. However, Bayesian paradigm is capable of referring directly to the quantity of interest itself.

To summarize, Bayesian probability is a systematic approach to describe subjective uncertainty using the mathematical language of probability. To introduce the logic of estimation and inference with Bayesian probability in detail, the next section discusses the three compositions of Bayesian analysis: prior distribution, data distribution, and posterior distribution.

2) The three Components of Bayesian Analysis: Prior distribution, Data distribution, and Posterior distribution

The most significant difference between Bayesian statisticians and traditional statisticians lies in the way they treat the population parameter θ . In Bayesian analysis, probabilities of θ are distributed in an infinite range. This means that Bayesian statisticians take parameter θ as a random variable, while observed data y is treated fixed. This is the exact opposite to the conventional thinking of the population parameter, because, the traditional statisticians think of sampled data as random variable and assume the population parameters to be fixed. Bayesian statistics believe that population parameter exist only as a subjective tool to help understand the underlying objectivity of the world (Park, 2013).

Every Bayesian analysis consist of three types of probability distributions. First is the probability distribution of data denoted as $P(y|\theta)$. Given a specific model of the parameters, researchers calculate how probable the data that they acquired are. Probability distribution of the data is identical to the likelihood function from the traditional perspective, but has a different connotation. Data distribution is a function that can have any value between 0 and 1, and the integral of the function is 1. Conversely, likelihood is just a constant with a specific value. It is impossible to articulate the probability distribution of the data that the researcher has gained with the frequentist paradigm.

The second element consisting the Bayesian analysis is the prior probability distribution. Prior distribution, $P(\theta)$, is the probability of all parameters prior to any data being collected. The reason that Bayesian analysis is liberating — allow direct statements about the probability of parameters of interest — is because researchers choose a prior probability distribution for the parameters before any investigation even begins. Acknowledging the fact that science is about objectivity, the sense of utilizing sources other than objective data might seem unusual or out of reasoning. Substantial debate on this issue has been going on for centuries (Berger, 2006; Kass and Wasserman, 1996). However, Bayesian analysts justify the arbitrariness of prior distributions by using (1) empirical prior distribution, (2) conjugate prior distribution, and (3) diffuse prior distribution. These concepts will be discussed in detail later in the paper.

The last component of the Bayesian analysis is the posterior distribution. The posterior distribution is the probability of parameters θ given observed data y . This distribution is the

final result that the researcher wants to know. Posterior distribution is valuable in that it reflects not only the data from the research but also prior judgments of the researcher. Finally, the power of Bayesian analysis comes from the posterior distribution, because it allows hypothesis testing, counterfactual simulation and forecasting.

3) The Bayes' Theorem

Bayes' theorem is the most systematic way to update the unknown fact with known information. With conditional probability, Bayes' theorem updates the probability (information) in question by using three types of distributions discussed above.

When our parameter in question, θ , is continuous random variable, the prior and posterior density distribution can be represented as density functions. In other words, Bayes' rule estimates $f(\theta|y)$. It calculates the probability of parameters in θ given sample y , representing the posterior probability function.

Thus, Bayesian estimation $f(\theta|y)$ can be written as:

$$f(\theta|y) = \frac{f(\theta, y)}{f(y)} \quad \text{where } f(\theta, y) = f(y|\theta) * f(\theta) \quad \text{and}$$

$$f(y) = \int_{-\infty}^{\infty} f(\theta, y) d\theta = \int_{-\infty}^{\infty} f(y|\theta) * f(\theta) d\theta$$

We can summarize the Bayes' theorem by formalizing the posterior density distribution as follows:

$$f(\theta|y) = \frac{f(y|\theta) * f(\theta)}{\int_{-\infty}^{\infty} f(y|\theta) * f(\theta) d\theta}$$

From this formula, posterior probability of parameter given the data are proportional to the probability of the data as informed by the parameters (sample distribution) multiplied by the prior probability of the parameter.

4) Computing posterior distribution

To illustrate how the Bayes' rule could be apply in the real world, example provided by Winkler (1972) will be introduced. Let's suppose that θ represents the market share of a new brand of a specific type of product. Since the product of this new brand is significantly different from other brands, we are uncertain about how much share this nouveau goods will gain from the market. However, we do know that if this product meets the demand well, it can lock in customers and attract virtually the entire market (meaning, θ might be close to 1). At the same time we also know that the opposite situation could happen (that is θ might be close to 0). It seems logical that θ is continuous and has a varying value between 0 and 1. Furthermore, we can assume that low values of θ are more probable than high values. We can thus define the prior distribution as follows:

$$f(\theta) = \begin{cases} 2(1-\theta) & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

In order to obtain more information on θ , we can collect samples. Let's say five consumers have purchased a new product in question; one purchased the new brand, the other four purchased other brands. The act of purchasing a new product is mutually exclusive and the outcome of the act is represented either as 'buy' or 'not buy'. This process can satisfactorily be a Bernoulli process. The probability that a randomly picked customer would buy the new brand will be the same as θ , the market share of the new brand. Keeping this fact in mind, we can effectively argue that

$$f(y|\theta) = \binom{5}{1} \theta(1-\theta)^4 = 5\theta(1-\theta)^4 .$$

Applying the above equation to the Bayes' theorem, we can acquire the posterior distribution of θ .

$$\begin{aligned} f(\theta|y) &= \frac{f(y|\theta) * f(\theta)}{\int_{-\infty}^{\infty} f(y|\theta) * f(\theta) d\theta} = \frac{5\theta(1-\theta)^4 * 2(1-\theta)}{\int_0^1 5\theta(1-\theta)^4 * 2(1-\theta) d\theta} \\ &= \begin{cases} 42\theta(1-\theta)^5 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Although the Bayesian statistics is a convenient method to revise probability distribution in respect to sample information, in reality, it is not as convenient as one might want to believe. The calculation required for the posterior distribution is too complex. Because of the difficulties in the application of Bayes' theorem, Bayesian statisticians have developed three ways to unburden this issue and they will be discussed below.

2. Bayesian Estimation and Inference

1) Empirical, Conjugate, and Diffuse prior distribution(de Finetti's Theorem)

In section IV.1.2) we talked about how the usage of prior probability distribution could be justified and in section IV.1.4) we discussed the complexity of calculating and implementing the Bayes' rule. However, thorough elaboration on these issues were procrastinated. There is a justifiable intent for such a delay. Although the issue raised in IV.1.2) and IV.1.4) might seem distinct, they are actually tightly knit together.

Statisticians endeavored to relieve the difficulty of application of Bayes' theorem. To begin with, empirical prior distribution is a prior consisted of historical data or is a distribution that is estimated from the data set itself. Empirical priors are advantageous in that it uses all observed data to estimate parameters that are associated with only a partial set of the data, as in multilevel modeling, where all of the data are used to estimate group's mean (Raudenbush and Bryk, 2002; Zyphur and Oswald, 2013). However, the use of the empirical prior is troublesome in that the observed data for estimation is redundant; the influence of the data is percolated in both the prior and sample distribution, causing the sample size to look twice as big as it actually is. The result is posterior probability being too narrow. Although some researchers used empirical priors in their studies (Efron, 2010), the usage of empirical priors are often at odds with the essence of Bayesian estimation, where priors are to be updated with new data.

Furthermore, Bayesian statisticians have developed the concept of conjugate prior distributions to alleviate the difficulty of applying the equation of posterior distribution. Basically conjugate prior distributions are families of distributions that are used to ease the computational burden. Although, posterior distribution is influenced by both sample and prior distribution, it is known that, once certain assumptions are made about the population

Table 2. Most widely used Exponential Family Forms and Conjugate Priors

Likelihood form	Conjugate Prior Distribution
Bernoulli	Beta
Binomial	Beta
Multinomial	Dirichlet
Negative Binomial	Beta
Poisson	Gamma
Exponential	Gamma
Gamma	Gamma
Normal for μ	Normal
Normal for σ^2	Inverse Gamma
Pareto for α	Gamma
Pareto for β	Pareto
Uniform	Pareto

that is being sampled, the data distribution is determined according to the chosen statistical assumption or model (Winkler, 1972). Simply put, if we can specify a particular data-generating model, the sample distribution can be known. After we determine a conjugate model, it is not so overwhelming to combine these family of distribution to get the posterior distribution.

Three properties are suggested as desirable properties for conjugate families: mathematical tractability, richness, and ease of interpretation. A prior distribution is mathematically tractable when it (1) is fairly easy to specify the posterior distribution given the prior distribution and sample distribution, (2) results in a posterior distribution that is also a member of the same conjugate family and (3) is feasible to calculate expectations from the prior distribution. Mathematical tractability is the most crucial reason for using conjugacy because it relieves the computational burden associated in Bayesian statistics. Moving on, a conjugate family of distributions should include distributions with various dispersion and shapes to be able to represent a wide variety of states of prior information. Without this property of richness there would not be sufficient number of distributions to accurately reflect ones prior information. Below is the most widely used conjugate family in current research from various fields. Lastly, specifying a conjugate family so that the researcher can readily interpret the prior information is necessary.

Conjugating prior distribution for the normal distribution for μ will be illustrated. The example is set as normal distribution for μ because this distribution is one of the most important and widely used distribution.

Suppose that a researcher is sampling from a normal population. We assume that he knows the variance of the population, σ^2 but not the mean of the population, μ . Although he does not know the exact value of the mean, he still has some prior information concerning the mean of the population. If his prior distribution of the uncertain quantity μ follows a normal distribution, the posterior distribution will also follow a normal distribution. Suppose that the prior distribution follows the form as below.

$$f'(\mu) = \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-(\mu-m')^2/\sigma'^2}$$

The above equation is a normal density function with mean m' and variance σ'^2 . If the researcher takes sample of size n and observes the sample mean of m , the posterior density is a normal density,

$$f''(\mu | y) = \frac{1}{\sqrt{2\pi\sigma''^2}} e^{-(\mu-m'')^2/\sigma''^2} \quad \text{where } y \text{ is the sample results.}$$

The posterior parameters m'' and σ''^2 can be determined from formulas

$$\frac{1}{\sigma''^2} = \frac{1}{\sigma'^2} + \frac{n}{\sigma^2} \quad \text{and} \quad m'' = \frac{\frac{1}{\sigma'^2} * m' + \frac{n}{\sigma^2} * m}{\frac{1}{\sigma'^2} + \frac{n}{\sigma^2}}.$$

n is the sample size and m is the sample mean. For the family of normal distributions, mathematical tractability is satisfied. Though the computations for posterior parameters seem complicated, it actually does not pose any serious trouble in calculating the parameters. The posterior distribution is a member of the same family as the prior distribution, so repeated applications of Bayes' theorem create no difficulties.

By transforming one of the parameters used to define the distribution, we can interpret the meaning of normal prior distribution. Consider the parameter n' the value of variance of the population in question divided by variance of sample mean.

$$n' = \frac{\sigma^2}{\sigma'^2}$$

This equation can be transformed into $\sigma'^2 = \frac{\sigma^2}{n'}$ so that the prior variance can be expressed in terms of n' and the population variance σ^2 . The prior distribution is thus a normal distribution with mean m' and variance $\frac{\sigma^2}{n'}$. Now if we do the same procedure with n'' with definition of

$$n'' = \frac{\sigma^2}{\sigma''^2}$$

it is possible to effectively show that $n'' = n' + n$ to make $m'' = \frac{n'm' + nm}{n' + n}$. Noticeably with the transformation of parameters of prior distribution (posterior distribution) is expressed in terms of n' and m' (n'' and m'').

What does this new parameter mean? By investigating $\sigma'^2 = \frac{\sigma^2}{n'}$ we can interpret n' as the sample size required to produce a variance of σ'^2 for a sample mean. This interpretation suggest that prior distribution is roughly equivalent to the information contained in a sample size of n' with sample mean m' . Under the influence of this interpretation, $n'' (= n' + n)$ and $m'' (= \frac{n'm' + nm}{n' + n})$ can be thought of as formulas for pooling the information from two samples. Posterior sample size is equal to the sum of prior sample size and data sample size. Also, the posterior sample mean is a weighted average of prior mean and data mean. We can safely conclude that in estimating parameter θ , the one with the larger sample (distribution with more information) receives more weight in the determination of θ .

To defend the skepticism of using prior distributions in estimation, some statisticians have used diffuse distribution. Suppose that a researcher wants to assess prior distribution in an environment with little or no information. Graphically, this situation can be represented as the Figure 2.

Let θ represent the parameter in question and y represent the sample result. Relative to the sample distribution, the prior distribution is flat. A flat distribution signifies that it does not contain much information and that it could be approximated by a constant function, namely $f'(\theta) = k$. By now, we know the fact that posterior probability function is proportional to the product of the prior density and the sample density function. However, since the prior

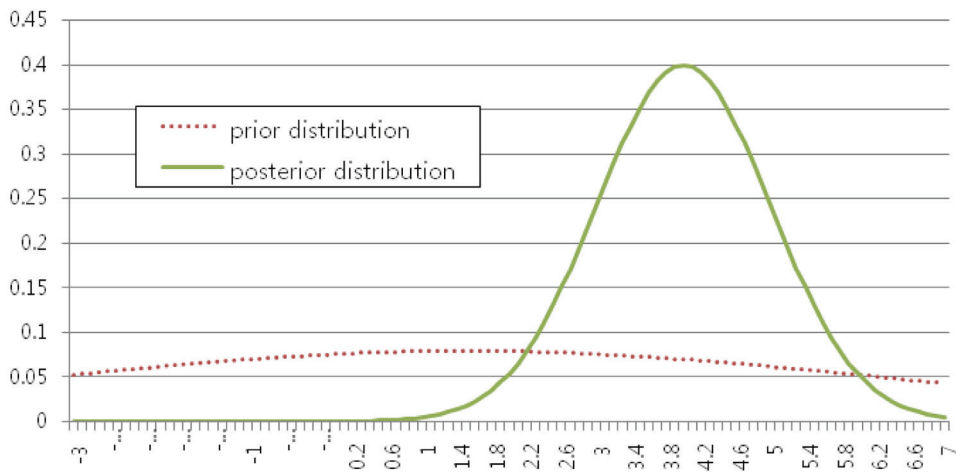


Figure 2. Prior and Posterior distribution

distribution contain no information and is expressed as a constant, we conclude that the posterior density is proportional to the sample function.

$$f''(\theta|y) \propto f(y|\theta)$$

This expression formalizes what was previously explained. If the prior distribution is relatively diffused to the sample distribution, the posterior distribution is dependent mostly on the sample distribution function. And this use of diffuse distribution as a prior distribution was one of the arguments for the Bayesian statisticians.

For Bayesian estimation, prior distribution is crucial. Our paper suggests three ways of how the subjectivity of distribution could be specified. De Finetti’s theorem advocates for the usage of prior functions in empirical studies. It essentially contends that exchangeable observations are conditionally independent given some latent variable to which an epistemic probability distribution would then be assigned. Thus, for researchers to analyze probabilistic question in a consistent and systematic way, they are required to utilize their prior knowledge or data (in the form of prior distribution).

However, we should note some criticisms to the usage of Bayesian statistics; Bayesian is not the answer to all of the research question. Bayesian statistics allows for too much variability

about the data and therefore is not suitable in considerations that have no adaptive relevance (Browser and Davis, 2012).

2) Markov Chain Monte Carlo: the MCMC revolution

The ultimate goal of Bayesian estimation is finding the posterior distribution of the population parameter by using sample distribution and prior distribution (Park, 2013). However, when a model encompasses more than two population parameters, Bayesian estimation requires multidimensional integral that makes the calculation of the posterior distribution very complicated. This complexity is the crucial constraint in the wide diffusion of Bayesian statistics.

The revolution of Bayesian statistic involves Markov chain Monte Carlo (MCMC) estimation (Efron, 2011). The most common method for Bayesian estimation is MCMC that allows specifying many types of priors. MCMC is an iterative process, where a prior distribution is specified and posterior values for each parameter are estimated in continuous iteration that form a 'chain'. Posterior values are estimated and these values draw the posterior distribution. MCMC starts from at least two starting points to show the convergence of the iteration process on a stable estimate of posteriors. We can understand the principle of MCMC most easily by using discrete Markov chain.

Markov chain is a time series random variable that is affected only by the adjacent time or state. In simple terms, a probability distribution of the next state is only depended on the current state and not on the events that precedes it. For example, let's assume that corrupted behavior in an organization follows such Markov chain below (example adapted from Park, 2013).

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.1 & 0.7 \\ 0.3 & 0.2 & 0.5 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

a_{ij} denotes the probability of current state i changing into state j in the next period. 1 stands for corrupted state, 2 for semi-corrupted state (since this is an example, let's not bother with philosophical definition of semi-corrupted state), and 3 for not corrupted state. For example,

a_{12} stands for the probability of currently corrupted organization changing into a semi-corrupted organization. Matrix A assumes that value to be 0.1. The fact that a_{33} has the highest value of 0.8 suggests that once an organization is in a clean state, it is highly likely that it will remain that way.

Stochastic matrix A stabilizes after a several iterations, regardless of any starting values. In other words, by using this stochastic matrix we can figure out the probability of a state of a randomly picked organization (whether it is corrupted, semi-corrupted or not corrupted). Let's look at how the distinct probability distributions changes after 20 steps of iteration.

$$A = \begin{pmatrix} 0.2 & 0.1 & 0.7 \\ 0.3 & 0.2 & 0.5 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

$$A^2 = A * A = \begin{pmatrix} 0.14 & 0.11 & 0.75 \\ 0.17 & 0.12 & 0.71 \\ 0.13 & 0.11 & 0.76 \end{pmatrix}$$

⋮

$$A^{20} = A^{19} * A = \begin{pmatrix} 0.1358025 & 0.1111111 & 0.7530864 \\ 0.1358025 & 0.1111111 & 0.7530864 \\ 0.1358025 & 0.1111111 & 0.7530864 \end{pmatrix}$$

A^{20} is the stochastic matrix after 20 iteration. If we multiply different initial distribution values, we can find out that the probability after 20 iteration does not change.

$$p_1=(0.2 \ 0.2 \ 0.6), p_2=(0.1 \ 0.5 \ 0.4), p_3=(0.3 \ 0.7 \ 0) \text{ then,}$$

$$p_1 * A^{20} = p_2 * A^{20} = p_3 * A^{20} = (0.1358025 \ 0.1111111 \ 0.7530864)$$

Although Markov chain starts with different values (p_1 , p_2 , and p_3) after certain amount of iteration, the influence of initial distribution fades out, stabilizing to a certain probability distribution.

In summary, MCMC uses computer simulation to figure out the posterior distribution that

complies the characteristic of Markov chain. MCMC is special because it is not sensitive to the initial values or the order of the simulation and it is destined to arrive at an equilibrium.

V. Investigating organizational science literature for use of Bayesian for NHST

Although classical frequentist prevail in current social science research, there has been an explosion of Bayesian application for the past 10 years. To illustrate, Journal of management issued a special session(2015, 41(2)) to meet the increasing demands of researchers. Through the online catalogue, Business Source Complete and Sage Journals, we searched related journals to check how scholars applied Bayesian approach to their researches. By searching the keyword “Bayesian” from 2005 to 2015, we have found 3,605 results. Since our interest lies specifically on the field of organizational studies, we narrowed the results to the ones that belong to the SSCI(Social Sciences Citation Index). The specification produced only 69 results.

Among the 69, 32 articles were about the methodology and the rest 37 were on empirical findings applying the Bayesian approach in their analysis. We have found out that almost half the portion of the published papers are simple guidance for researchers just beginning to learn about Bayesian and concluded that the field of management still needs substantive development in the application of Bayesian to the research.

Until now, the use of Bayesian approach is predominantly restricted to SEM as a Bayesian information criterion (BIC) to quantify a model’s goodness of fit to data. BIC is easily transformed into an estimate of the Bayes factor.

Table 3. Articles related to Bayesian approach (2007 – 2016)

	Bayesian referred	Methodological issues	Empirical paper
Articles*	69	32 (46.4%)	37 (53.6%)

Note. * published from 2007–2016, including Journal of Management, Journal of Organizational Behavior, Journal of Applied Psychology, Academy of Management etc.

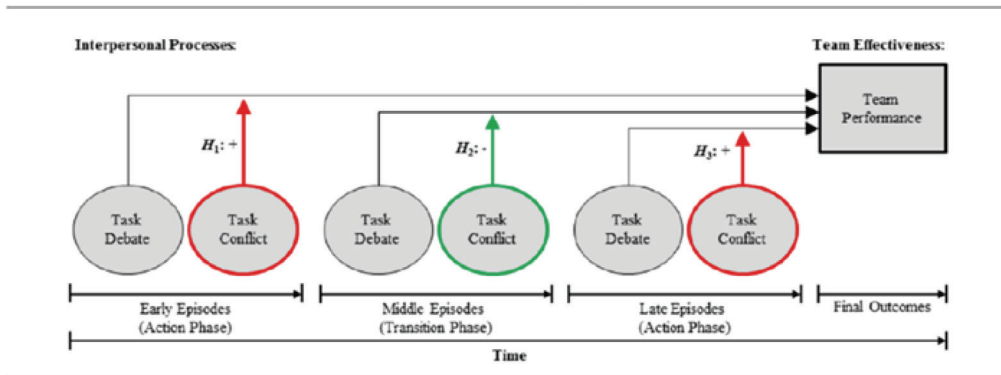


Figure 3. Research Question and Conceptual Model (Johnson, van de Schoot, Delmar and Crano, 2015)

Among the empirical papers (37 articles), BSEM (Bayesian Structural Equation modeling) is the most prevalent method. The goal of BSEM is to identify a good model that minimizes the discrepancy between predicted outcome and actual value in a parsimonious way and maximizes the generalizability to new samples (Reise et al., 2000). To fulfill this goal, Stromeier et al. (2005) recommended to use not only frequentist criterion such as RMSEA, but also Bayesian information criterion. As we have discussed earlier, the more accurate the prior distribution, the more likely that the model becomes parsimonious.

We continue by introducing contexts that would be useful to use Bayesian framework. A BIC statistics is especially useful for complex model including many latent and categorical variables. It stems from comparison of pairwise Bayes factors. The BIC statistics allow one to successfully compare the fitness among numerous models. Since the BIC introduces a penalty term for the number of parameters in the model, the problem of overfitting can be resolved.

Johnson, van de Schoot, Delmar and Crano (2015) used Bayesian estimation in multi-level structural equation modeling. In their research, they tried to explain why dissenting in-group minorities could improve team performance.

In the study, the authors implemented the Bayesian approach in two ways. First, by incorporating Bayesian estimation, they tried to overcome the limitation of maximum likelihood estimation (MLE) such as the assumption of normality. Specifically, they assume the moderating effect between task conflict and task debate. The traditional method had had

a problem in that the product terms are known to have non-normal distributions (Shrout and Bolger, 2002). However, since Bayesian estimation is not bound by the normality assumptions, the problem caused by product terms could be resolved. Secondly, they used informative hypotheses and Bayesian model selection to test the hypotheses directly. By using Bayes factor, they selected the best fitting model. In the research, they elaborated the interpersonal dynamics longitudinally. Classical hypothesis testing can evaluate a single parameter and a sequence of hypothesis to compare each other. However, when three issues related to one another and interact sequentially, it's better to use Bayes factor because it allows researchers to evaluate their expectation directly against one another.

$$H_4 : H_1, H_2, H_3$$

$$H_4 : (\hat{Y}_2 > \hat{Y}_1), (\hat{Y}_4 < \hat{Y}_3), (\hat{Y}_6 > \hat{Y}_5)$$

In a case such as H_4 , which synthesizes H_1, H_2, H_3 , traditional NHST requires a tremendous number of significance tests that are not necessarily independent, three tests are needed for each individual effect, then multiple pairwise comparison tests of the effects need to be conducted to determine if they are significantly different from each other and in the expected direction. Subsequently, it will be very difficult to combine the resulting set of p-values into one single answer to the question of whether the hypothesis is supported by the data. Also, multiple testing is no direct way to understand what the null hypothesis is in this context and make it unclear which p-value is counted for conclusion. It fails to immediately assess the research question of interest.

Sang Eun Woo et al. (2016) tried to delineate how various personality traits (bright vs. dark traits) are associated with turnover criterion variables (i.e., turnover speed and reasons). To elaborate their reasoning, they used the Bayesian competing risks survival analysis. To investigate turnover risk, most researchers use estimated proportional hazards rate models, commonly referred to as cox regression model. It assumes that changes in the independent variables produce proportional changes in the baseline hazard rate. This survival method is commonly accepted because that it allows examination of both why and when individuals leave their organizations. However, researchers indicated the limitation of previous survival

analyses. They are usually coded turnover as a dichotomous event, which does not provide insight into how various factors relate to specific turnover reasons. So, the researchers chose a competing risks approach models which can provide fuller explanation. By using Bayesian approach, researchers estimated the survival and hazard functions which are essential aspects of any survival analysis because they provide information of time-varying predictors and time-dependent effects. Bayesian estimation served important advantages. First, by using Bayesian basic concept to utilize the priori, the model can contain important previous findings and complex directional effects. (i.e., organizational commitment predicts negative relations for turnover speed, Learning predicts positive for turnover speed etc.) Second, using a Bayesian approach to produce a posterior probability is advantageous for selecting more precise model because it shows direct probability that true value can be contained in confidence intervals. To grasp incremental changes of predictors over time like this model, it is good for Bayesian approach to provide both point estimates such as mean and median of the distribution and intervals that contain probability distribution. Third, by using Bayes factor, it can lead to a better selection for competing models with deviance information criterion (DIC). With help of Bayesian approach, the authors can effectively induce the conclusion that personality traits predicts timing and reasons of turnover and they can be diverged depends on timing.

Cordery et al. (2015) clarified the role of communities of practices (CoPs) for organizational changes by facilitating the transfer of best practices. To delineate the process, the authors used a Bayesian change point detection model to estimate for the change probability linked in the adoption of CoPs for better and improved practices. They argued the Bayesian analysis provides fruitful implications related to organizational interventions. As we previously discussed, Bayesian approach can evaluate the complex change with direct estimation for timing of change and point of change as distinguished parameters. The success of study depends on how precisely grasp the difference states before and after the change. For frequentist way, it's almost impossible for making null hypothesis for 5-year-long CoPs variation and changes. In contrast, Bayesian approach provides the posterior probability that change occurs at each time point in the observational period. It can make inference for prior complex probability before the change. With this methodology, they turned out that CoPs are autonomously generates and screens the operational procedures beneficial for the company without the institutional interventions even

if it need some time to be embedded.

In summary, based on representative examples, we can conclude Bayesian approach is especially powerful for the complex directional relations among variables or time-variant effects assumed in a model.

Although, this section provides the starting point to further research on the use of Bayesian approach in organizational science, there are some limitations. By keyword search “Bayesian”, it’s possible to exclude the articles without mention “Bayesian”. Further analysis needs to be accomplished to look up the frequency as to which the Bayesian approach may better explain organizational science which could provide additional insights. In line with explosion of interdisciplinary studies, it would be beneficial to examine more wide range of management literature to determine when the Bayesian approach adds the value most.

VI. Conclusion

Traditionally, the NHST has been the most widely used method of analysis. However, it has caused numerous controversies regarding the validity of extracting information from data and of guiding the formation of scientific conclusions. Acknowledging the shortcomings of NHST, we introduced several ways to complement the traditional approach, including the use of effect size, confidence interval and Bayesian statistics. Among those the Bayes approach was strongly focused.

In the paper we tried to elaborate the powerfulness and potential Bayes theorem had. Especially, Bayesian statistics is competent because it calculates the probability of the actual question of interest and this probability is more communicable to the audience as a whole. There are vulnerabilities to the usage of Bayesian, however, the implementation of the analysis is surprisingly disappointing. We believe that the under-use of this approach is due to inertia and lack of knowledge.

There is no one universal way to rational inference. The role of researcher is to infer scientifically by using methodology most well suited to research questions. Thus, we are confident that the Bayesian approach would open the new door to the field of organization

science.

References

- Abelson, R. P. (1995), *Statistics as Principled Argument*, Hillsdale, NJ: Erlbaum.
- American Psychological Association (1994), *Publication Manual of the American Psychological Association* (4th ed.), Washington, DC: Author.
- Asher, W. (1993), "The Role of Statistics in Research," *Journal of Experimental Education*, 61, 388-393.
- Berger, J. O. (2006), "The Case for Objective Bayesian Analysis," *Bayesian Analysis*, 3, 385-402.
- Bonett, D. G. and T. A. Wright (2007), "Comments and Recommendations Regarding the Hypothesis Testing Controversy," *Journal of Organizational Behavior*, 28(6), 647-659.
- Bower, J. S. and C. J. Davis (2012), "Bayesian Just-so Stories in Psychology and Neuroscience," *Psychological Bulletin*, 138(3), 389-414.
- Breaugh, J. A. (2003), "Effect Size Estimation: Factors to consider and mistakes to avoid," *Journal of Management*, 29(1), 79-97.
- Carlson, K. D. and F. X. Ji (2011), "Citing and Building on Meta-analytic Findings: A review and recommendations," *Organizational Research Methods*, 14(4), 696-717.
- Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences* (Rev. ed.), NY: Academic Press.
- Cordery, John L., Edward Gibson, Cristina B. Cripps, and John E. Mathieu (2015), "The Operational Impact of Organizational Communities of Practice," *Journal of Management*, 41(2), 644-664.
- Cox, D. R. and D. V. Hinkley (1979), *Theoretical Statistics*, London, England: Chapman and Hall.
- Cronbach, L. J. and R. E. Snow (1977), *Aptitudes and Instructional Methods: A handbook for research on interactions*, NY: Irvington.
- Crow, E. L. (1991), Response to Rosenthal's comment "How are we doing in soft psychology?" *American Psychologist*, 46, 1083.
- Cumming, G., J. Williams, and F. Fidler (2004), "Replication and Researchers' Understanding of Confidence Intervals and Standard Error Bars," *Understanding Statistics*, 3(4), 299-311.
- Dalton, D. R. and C. M. Dalton (2008), "Meta-analyses: Some very good steps toward a bit longer journey," *Organizational Research Methods*, 11(1), 127-147.

- Daston, L. (1994), "How Probabilities Came to be Subjective and Objective," *Historia Mathematica*, 21, 330-344.
- Duhachek, A. and D. Iacobucci (2004), "Alpha's Standard Error (ASE): An accurate and precise confidence interval estimate," *Journal of Applied Psychology*, 89(5), 792-808.
- Eagle, A. (2004), "Twenty-one Arguments Against Propensity Analyses of Probability," *Erkenntnis*, 60, 371-416.
- Edwards, J. R. (2008), "To Prosper, Organizational Psychology should... Overcome Methodological Barriers to Progress," *Journal of Organizational Behavior*, 29, 469-491.
- Edwards, J. R. and J. W. Berry (2010), "The Presence of Something or the Absence of Nothing: Increasing theoretical precision in management research," *Organizational Research Methods*, 13(4), 668-689.
- Efron, B. (2010), *Large-scale Inference: Empirical Bayes methods for estimation, testing and prediction*, Cambridge, England: Cambridge University Press.
- Efron, B. (2011), "The Bootstrap and Markov-chain Monte Carlo," *Journal of Biopharmaceutical Statistics*, 21, 1052-1062.
- Fern, E. F. and K. B. Monroe (1996), "Effect-size Estimates: Issues and problem in interpretation," *Journal of Consumer Research*, 23(2), 89-105.
- Fisher, R. A. (1959), *Statistical Methods and Scientific Inference* (2nd ed.), NY: Hafner Publishing.
- Frick, R. W. (1996), "The Appropriate Use of Null Hypothesis Testing," *Psychological Methods*, 1, 379-390.
- Galavotti, M. C. (2005), *Philosophical Introduction to Probability*, Palo Alto, CA: CSLI.
- Gill, J. (1999), "The Insignificance of Null Hypothesis Significance Testing," *Political Research Quarterly*, 52(3), 647-674.
- Hagen, R. L. (1997), "In Praise of the Null Hypothesis Statistical Test," *American Psychologist*, 52, 15-24.
- Johnson, R. A., van de R. Schoot, F. Delmar, and D. W. Crano (2015), "Social Influence Interpretation of Interpersonal Processes and Team Performance Over Time Using Bayesian Model Selection," *Journal of Management*, 41(2), 574-606.
- Kass, R. E. and L. Wasserman (1996), "The Selection of Prior Distributions by Formal Rules," *Journal of the American Statistical Association*, 91, 1343-1370.
- Kileen, P. R. (2005), "An Alternative to Null-hypothesis Significance Tests," *Psychological Science*, 16(5), 345-353.
- Kim, M. (2011), "지리교육 연구에서의 양적 방법론: 효과 크기의 개념과 활용"[Quantitative Methods in Geography Education Research: Concept and application of effect size], *Korean*

- Association of Geographic and Environmental Education*, 19(2), 205-220.
- Kirk, R. E. (1996), "Practical Significance: A concept whose time has come," *Educational and Psychological Measurement*, 56, 746-759.
- Levine, T. R., R. Weber, C. Hullest, H. S. Park, and L. L. M. Lindsey (2008), "A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research," *Human Communication Research*, 34(2), 171-187.
- McGraw, K. O. (1991), Problems with BESD: A comment on Rosenthal's "How are we doing in soft psychology?" *American Psychologist*, 46, 1084-1086.
- Mckee, A. R. and C. C. Miller (2015), "Institutionalizing Bayesianism within the Organizational Sciences: A practical guide featuring comments from eminent scholars," *Journal of Management*, 41(2), 471-490.
- Meehl, P. E. (1967), "Theory-testing in Psychology and Physics: A methodological paradox," *Philosophy of Science*, 34, 103-115.
- Meehl, P. E. (1978), "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology," *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Nakagawa, S. and I. C. Cuthill (2007), "Effect Size, Confidence Interval and Statistical Significance: A practical guide for biologists," *Biological Reviews*, 82(4), 591-605.
- Nickerson, R. S. (2000), "Null Hypothesis Significance Testing: A review of an old and continuing controversy," *Psychological Methods*, 5(2), 241-301.
- Park, J.-H. (2014), "베이저안 사회과학방법론이란 무엇인가?" [What is Bayesian Methodology?], *Peace Studies*, 22(1), 481-529.
- Raudenbush, S. W. and A. S. Bryk (2002), *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.), Newbury Park, CA: Sage.
- Reise, S. P., N. G. Waller, and A. L. Comrey (2000), "Factor Analysis and Scale Revision," *Psychological Assessment*, 12, 287-297.
- Richardson, J. T. (1996), "Measures of Effect Size," *Behavioral Research Methods, Instruments, & Computers*, 28, 12-22.
- Rosenthal, R. (1991), *Meta-analytic Procedures for Social Research*, Newbury Park, CA: Sage.
- Rosenthal, R. (1994), "Parametric Measures of Effect Size," in H. Cooper and L. V. Hedges eds., *The handbook of research synthesis* (pp. 231-444), NY: Russell Sage.
- Rosnow, R. L. and R. Rosenthal (1989), "Statistical Procedures and the Justification of Knowledge in Psychological Science," *American Psychologist*, 44, 1276-1284.
- Rozeboom, W. W. (1960), "The Fallacy of the Null-hypothesis Significance Test," *Psychological Bulletin*, 57(5), 416-428.

- Savage, I. R. (1957), "Nonparametric Statistics," *Journal of the American Statistical Association*, 52, 331-344.
- Schmidt, F. L. (1996), "Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for training of researchers," *Psychological Methods*, 1(2), 115-129.
- Shrout, P. E. and N. Bolger (2002), "Mediation in Experimental and Non-experimental Studies: New procedures and recommendations," *Psychological Methods*, 7, 422-445.
- Stromeyer, William R., Jason W. Miller, Rajendran Sriramachandramurthy, and Richard DeMartino (2015), "The Prowess and Pitfalls of Bayesian Structural Equation Modeling," *Journal of Management*, 41(2), 491-520.
- Thompson, B. (2002), "What Future Quantitative Social Science Research could Look Like: Confidence intervals for effect sizes," *Educational Researcher*, 31, 25-32.
- Trafimow, D. (2014), "Editorial," *Basic and Applied Social Psychology*, 36(1), 1-2.
- Trafimow, D. and M. Marks (2015), "Editoria," *Basic and Applied Social Psychology*, 37, 1-2.
- Wagenmakers, E. J. (2007), "A Practical Solution to the Pervasive Problems of p values," *Psychonomic Bulletin & Review*, 14(5), 779-804.
- Winkler, R. L. (1972), *An Introduction to Bayesian Inference and Decision*, NY: Holt, Rinehart and Winston.
- Woo, S., M. Chae, A. Jebb, and Y. Kim (2016), "A Closer Look at the Personality: Turnover relationship," *Journal of Management*, 42(2), 357-385.
- Zyphur, M. J. and F. L. Oswald (2013), "Bayesian estimation and inference: A user's guide," *Journal of Management*, 41(2), 390-420.