EXAMINING SOME BAYES FACTORS ON THEIR DECISIONS FOR TESTING HYPOTHESIS FOR ONE WAY ANOVA WITH RANDOM EFFECTS

EGBURONU, O. D[1]., & ABIDOYE, A. O.[2]
[1]Department of Sciences, CLAPAI Orphanage High School, Jos
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria
E-mail: solatayo2003@gmail.com, abidoye@unilorin.edu.ng, solatayo2003@yahoo.com
Phone No: +234(0)8050663983;

Abstract
Bayesian statistics and in particular Bayes Factors have been proposed as an alternative to improve the scientific decision making when testing a hypothesis; this became necessary after the null hypothesis significant test based on p-values were criticized by researchers in various fields. A lot of Bayesian procedures have been proposed overtime but its wide spread adoption is scant. This article examined a prior sensitive Bayes Factors proposed by (Wang and Sun 2013) as well as a Bayesian Information Criterion-based Bayes factor proposed by (Faulkenberry 2018) on their decisions for testing the null hypothesis for One Way ANOVA with random effects. We illustrated the two procedures using simulated studies under two cases: Case 1: factor unit is fixed while observation per unit is increasing (i.e. random). Case 2: observation per unit is fixed while number of factor unit is increasing (i.e. random). The study revealed that in all the two cases, the two Bayes factors were consistent in increasing the weight of evidence in support of the null hypothesis of zero between factor variability; but as the sample sizes became large, the prior sensitive Bayes factor become impracticable. This impracticality situation was as a result of the Gamma function involved in its computational formular.

Keywords: Bayesian, Bayes Factor, Random effects, P-value, Analysis of Variance

Introduction

Assessing variability according to distinct factors in data is a fundamental technique of statistics, (Steven & Reinhard, 2013). Most researchers using Analysis of Variance (ANOVA) procedures choose a fixed-effects model, even though they may not realize that they are making this choice or realize its consequences. However, a random or mixed effects model may be a more appropriate fit for many research designs. The choice has implications for the generalizability of the findings, for the type of statistical questions that can be asked, for the fit between data and the model, and for the conceptual match between the model and the theory. From a Bayesian perspective, random factors are introduced to increase generalizability and accuracy. From a frequentist' point of view, random factors allow for more tightfisted models with fewer parameters. Without a good knowledge of the statistical foundations, it may be hard to determine which factors are best treated as random (Jansen, 2011).

If a fixed way is misclassified as random, it will be subject to an overly conservative test of statistical significance and therefore the likelihood of making a Type II error (not rejecting a false null hypothesis) will increase (Wike & Church, 1976). Inversely, if a random way is misclassified as fixed, there is a greater chance of making a Type I error (falsely rejecting a true null hypothesis), (Clark, 1973). In addition, if a random way is classified as fixed, the results of the study cannot be generalized beyond the levels that are utilized in the study, (Clark, 1973). To resolve the issues of concern above, (Gelman, 2005) suggested that all factors in the model should be treated as random.

Owing to its importance and simplicity, ANOVA is taught in virtually every applied statistics course. Nevertheless, the Bayesian hypothesis testing literature on ANOVA is scant. A Bayesian approach to test the hypothesis is to use Bayes factors comparing the hypothesis/models with

and without the random effects in question; Bayes factors have been advocated as superior to p-values for assessing statistical evidence in data. Despite the advantages of Bayes factors and the drawbacks of p-values, inference by p-values is still nearly ubiquitous. One impediment to adoption of Bayes factors is a lack of practical development, particularly a lack of ready-to-use formulas and algorithms.

Bayesian methods have become increasingly popular in almost all scientific disciplines, (Poirier, 2006). One important reason for this gain in popularity is the ease with which Bayesian methods can be applied to relatively complex problems involving, for instance, hierarchical modeling or the comparison between non nested models. However, Bayesian methods can also be applied in simpler statistical scenarios such as those that feature basic testing procedures. Prominent examples of such procedures include ANOVA and the Student *t*-test; these tests are the cornerstone of data analysis in fields such as biology, economics, sociology, and psychology, (Wetzel, Grasman & Wagenmakers, 2012).

Numerical values of population characteristics are typically expressed in terms of a parameter and numerical descriptions of the subset that make up a dataset. Before a dataset is obtained, the numerical values of both the population characteristics and the dataset are uncertain. After a dataset is obtained, the information it contains can be used to decrease our uncertainty about the population characteristics. Quantifying this change in uncertainty is the purpose of Bayesian inference. See Nathoo and Masson (2016). In Bayesian estimation, uncertainty about parameters is quantified by probability distributions.

Methodology

Bayes rule

Suppose, we have a model μ and we wish to estimate the model parameters (***θ)***. Then, we have to define a *prior distribution* over these parameters; $p(\theta|\mu)$. When data ***Y*** comes in, this prior distribution $p(\theta|\mu)$ is updated to yield the *posterior distribution* $p(\theta|Y,\mu)$.
According to Bayes' rule,

$$p(\theta / Y, \mu) = \frac{p(Y / \theta)\, p(\theta / \mu)}{p(Y)} \qquad (2.1)$$

$$= \frac{p(Y|\theta,\mu)p(\theta|\mu)}{\int p(Y|\theta,\mu)p(\theta|\mu)\, d\theta}$$

$$= \alpha \;\; p(Y / \theta, \mu)\, p(\theta / \mu) \qquad (2.2)$$

It is important to note that Bayes' rule does not tell us what our beliefs should be; it tells us how they should change after seeing new information, (Wetzel, Grasman and Wagenmakers 2012).

Bayes factor

Consider the balanced one-way analysis-of-variance (ANOVA) random effects model,

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1,2,...,k \quad \text{and} \quad j = 1,2,...,m \qquad (2.3)$$

In such a balanced variance components model (2.2), we are often interested in evaluating whether the random effects should be included, which is equivalent to testing the null hypothesis:

$$M_0 : \alpha_i = 0 \quad \text{against} \quad M_1 : \alpha_i \neq 0 \qquad (2.4)$$

Kass and Raftery (1995) defined the Bayes factor as "a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another." Simply stated, the Bayes factor is a number, a ratio of one model's odds to the odds of another model.

The Bayes factor for comparing $M_1$to M_0given by (2.3) can be written as:

$$BF_{10} = \frac{m_1(Y)}{m_0(Y)} \qquad (2.5)$$

where,

$$m_1(Y) = p(Y / m_1)$$

and

$$m_0(Y) = p(Y / m_0)$$

Hypothesis Testing

The decisions about the null hypothesis for conducting One Way ANOVA with random effects test was examined using two (2) Bayes factor namely:

(i) Prior Sensitive Bayes factor (Wang & Sun 2013).

(ii) BIC-based Bayes factor (Faulkenberry 2018).

The behaviours of the Bayes factors were examined using simulation studies under two cases, namely: Case 1 and Case 2.

CASE 1: Factor/treatment unit (k) is fixed while observations per units (m) are increasing.

CASE 2: Number of observations per units (m) is fixed while factor/treatment units (k) are increasing. All the methodologies are discussed below.

Simulation Study:

Data sets were simulated using the native functions implemented in the R software for statistical computing (version 3.4.0 for Windows, R Core Team, 2017)from a standard normal population$N(\mu = 0, \sigma = 1)$. Simulation was generated using random seed sets to simplify replication.

Bayes Factor for One Way Random Effect Model
(Proposed by Wang and Sun (2013)

Consider the balanced one-way analysis-of-variance (ANOVA) random effects model,

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1,2,...,k \text{ and } j = 1,2,...,m \qquad (2.6)$$

Where, y_{ij} is the j[th] observation associated with the unit i and μ represents the unknown intercept. Here $k\ (\geq 2)$ is the number of factor/treatment units and $m\ (\geq 2)$ is the number of observations per unit. It is assumed that the random effect (α_i) and the error term (ε_{ij}) are mutually independent, and that $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ for all i and j, The unknown parameters $(\sigma_\alpha^2 \; and \; \sigma^2)$ are called variance components.

The Wang and Sun (2013) Bayes factor for obtaining the weight of evidence in support of the null hypothesis is given by

$$BF_{01} = \frac{\Gamma\left(\frac{mk-1}{2}\right)\Gamma(\alpha+1)}{\Gamma\left(\frac{k}{2}+\alpha+\frac{1}{2}\right)\Gamma\left(\frac{mk-k}{2}\right)}\left(\frac{SSE}{SST}\right)^{\frac{(mk-k-2)}{2+\alpha}} \quad (2.7)$$

where

$SSE = \sum_{i=1}^{k}\sum_{j=1}^{m}\left(y_{ij\cdot} - \bar{y}_{i\cdot}\right)^2$ is the sum of square error (2.8)

$SST = \sum_{i=1}^{k}\sum_{j=1}^{m}\left(y_{ij} - \bar{y}_{\cdot\cdot}\right)^2$ is the sum of square total (2.9)

Wang and Sun (2013) has established through simulation studies that the Bayes factor equation (2.7) is robust to a choice of $\alpha \in \left[-\frac{1}{2},\ 0\right]$.

The Bayes factor above has an explicit closed form expression without integral representation. This can be easily calculated using statistical packages, (Wang and Sun 2013). Based on work by (Raftery 1995) and (Wagenmakers 2007), (Faulkenberry 2018) demonstrated a method for estimating Bayes factors using the BIC. The (Faulkenberry 2018) BIC-based Bayes factor for obtaining the weight of evidence in support of the null hypothesis is given by

$$BF_{01} = \sqrt{nf_1\left(\frac{Fdf_1}{df_2}+1\right)^{-n}} \quad (2.10)$$

where

n is the total sample size, F is the F-statistic obtained from the ANOVA table, df_1 is the first degrees of freedom (k-1) and df_2 is the second degrees of freedom (mk-k).

Note that since $BF_{10} = \frac{1}{BF_{01}}$, the formula can be used flexibly to assess evidence for either the null hypothesis or the alternative hypothesis, depending on the researcher's needs. Conclusion on the Bayes factor values is drawn from table provided by Raftery (1995).

Table 1: Decision Rule Table For Bayes Factor Interpretation

Bayes Factor (BF_{01})	**Evidence For the null hypothesis (H_0)**
$1-3$	Not worth more a mere mention
$3-10$	Substantial
$10-100$	Strong
$>\ 100$	Decisive

Source: Raftery (1995)

The study will be carried out in the following steps:

Steps:

(i) Each case (1 and 2) will be studied under five (5) sub-cases namely (A, B, C, D and E) corresponding to the different number of factors (**k**)/observations (**m**).

(ii) For each sub-case (A, B, C, D and E), data will be simulated for the set **k** and **m** combination.

(iii) The frequentist One Way ANOVA table summary will be computed using the simulated data for each sub-case.

(iv) The Wang and Sun (2013) Prior Sensitive Bayes factor will be computed using results in (iii) above.

(v) Faulkenberry (2018) BIC-based Bayes factor will be computed will be computed using results in (iii) above.

These five steps will be replicated for all the sub-cases under the **case 1** and **case 2.**

Results and Discussion

CASE 1: Factor/Treatment Unit (K) Is Fixed While Number of Observations Per Units (M) Is Increasing

The factors/treatments are fixed at $(\boldsymbol{k = 5})$ units. (i.e the factors will be Factor A, B, C, D and E), whereas the observations per units (m) will be increasing. Five (5) different sample sizes per unit were simulated $(\boldsymbol{m = 5, 10, 20, 50\ and\ 100})$. Simulations were generated using random seeds to simplify replication. The methodologies discussed above are illustrated below using **Case 1A** $(k = 5\ and\ m = 5)$.

CASE 1A $(\boldsymbol{k = 5\ and\ m = 5})$:

Table 1: Simulated data for **CASE 1A $(\boldsymbol{k = 5\ and\ m = 5})$**

Factors/Treatment Units (k)	Observations per factor (m)				
A	-0.90	0.18	1.59	-1.13	-0.08
B	0.13	0.71	-0.24	1.98	-0.14
C	0.42	0.98	-0.39	-1.04	1.78
D	-2.31	0.88	0.04	1.01	0.43
E	2.09	-1.20	1.59	1.95	0.00

$$SST = \sum_{i=1}^{5}\sum_{j=1}^{5}(y_{ij} - \bar{y}_{..})^2 = 31.3915$$

$$SSB = m\sum_{i=1}^{5}(\bar{y}_{i.} - \bar{y}_{..})^2 = 2.9763 \qquad MSB = \frac{SSB}{k-1} = 0.7441$$

$$SSE = SST - SSB = 28.4152 \qquad MSE = \frac{SSE}{m(k-1)} = 1.4208$$

For the frequentist ANOVA, we seek to test the hypothesis,

$H_0 : \mu_1 = \mu_2 = ... = \mu_{10}$ against $H_1 : \mu_i \neq \mu_j$ for $i \neq j$ (3.1)

Table 2: Summary of One Way ANOVA table for Case 1A $(\boldsymbol{k = 5\ and\ m = 5})$

Source of Variation	Degree of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F-Ratio	p-value (p)
Between Groups	4	2.9763	0.7441	0.5237	0.7194
Within Groups (Error)	20	28.4152	1.4208		
Total	24	31.3915			

Frequentist' Decision rule: Reject H_0 if $P_{value} < significance\ level\ (\alpha = 0.05)$

Frequentist' Conclusion: Since P_{value} = 0.7194 > the significance level (α) = 0.05 and F-ratio = 0.5237 > the significance level (α) = 0.05 we do not reject the null hypothesis stated in equation (2.4). This implies that the five treatment means are the same.

Wang and Sun (2013) Bayes Factors for Case 1 A (k = 5 and m = 5):
We seek to test the hypothesis,

$$H_0 : \sigma_\alpha^2 = 0 \text{ against } H_1 : \sigma_\alpha^2 \neq 0 \quad (3.2)$$

The Wang and Sun (2013) Bayes factor for obtaining the weight of evidence in support of the null hypothesis for $(\boldsymbol{k = 5\ and\ m = 5})$ is computed as follows:

$$BF_{01} = \frac{\Gamma\left(\frac{mk-1}{2}\right)\Gamma(\alpha+1)}{\Gamma\left(\frac{k}{2}+\alpha+\frac{1}{2}\right)\Gamma\left(\frac{mk-k}{2}\right)}\left(\frac{SSE}{SST}\right)^{\frac{(mk-k-2)}{2+\alpha}}$$

$$= \frac{\Gamma\left(\frac{25-1}{2}\right)\Gamma\left(-\frac{1}{2}+1\right)}{\Gamma\left(\frac{5}{2}+\left(-\frac{1}{2}\right)+\frac{1}{2}\right)\Gamma\left(\frac{25-5}{2}\right)}\left(\frac{28.4152}{31.3915}\right)^{\frac{(25-5-2)}{2+\left(-\frac{1}{2}\right)}} = 34.04$$

The Bayes factor $BF_{01} = 34.04$, signifies that the data has a strong evidence in support of the null hypothesis of no variability between the five factors/treatments stated in equation (3.2). This can be seen in Table 3. Its inverse $BF_{10} = \frac{1}{34.04} = 0.029$ indicates negligible evidence that the data could occur under the alternative hypothesis where α = -0.5.

Table 3: Decision Rule Table For Bayes Factor Interpretation

Bayes Factor (BF_{01})	**Evidence For the null hypothesis (H_0)**
$1-3$	Not worth more a mere mention
$3-10$	Substantial
$10-100$	Strong
$>\ 100$	Decisive

Faulkenberry (2018) BIC-based Bayes Factor for Case 1A $(\boldsymbol{k = 5\ and\ m = 5})$:
We seek to test the hypothesis,

$$H_0 : \sigma_\alpha^2 = 0 \text{ against } H_1 : \sigma_\alpha^2 \neq 0 \quad (3.3)$$

The Faulkenberry (2018) BIC-based Bayes factor for obtaining the weight of evidence in support of the null hypothesis for $(\boldsymbol{k = 5\ and\ m = 5})$ is computed as follows:

$$BF_{01} = \sqrt{n^{df_1}\left(\frac{Fdf_1}{df_2}+1\right)^{-n}}$$

Where, n is the total sample size $(mk) = 25$. Hence,

$$BF_{01} = \sqrt{25^4\left(\frac{0.5237\times4}{20}+1\right)^{-25}}$$
$$= 179.94$$

The Bayes factors of $BF_{01} = 179.94$ indicates a decisive evidence in support of the null hypothesis of no variability between the five factors/treatments stated in equation (3.3). This can be seen in Table 3. Although, it's inverse $BF_{10} = \frac{1}{179.94} = 0.006$ indicates negligible evidence that the data will occur under the alternative hypothesis. This is also more informative than the frequentist conclusion would have offered because it provides information about the null and the alternative hypothesis.

Summary of Discussion for Case 1:
The methodologies illustrated in **case 1A** above were replicated for the four (4) other sub-cases under case 1 i.e

$\boldsymbol{k = 5\ and\ m = 10},\ \boldsymbol{k = 5\ and\ m = 20},\ \boldsymbol{k = 5\ and\ m = 50},\ \boldsymbol{k = 5\ and\ m = 100}$

The results are summarized in Table 4:

Table 4: Summary of the Bayes factors computations for Wang and Sun (2013) and Faulkenberry (2018) under Case 1

		$Case\ 1\ (k\ is\ fixed\ at\ 5\ while\ m\ is\ increasing)$				
	Formular	**m=5**	**m=10**	**m=20**	**m=50**	**m=100**
Wang and Sun (2013) (BF)	$BF_{01} = \frac{\Gamma\left(\frac{mk-1}{2}\right)\Gamma(\alpha+1)}{\Gamma\left(\frac{k}{2}+\alpha+\frac{1}{2}\right)\Gamma\left(\frac{mk-k}{2}\right)}\left(\frac{SSE}{SST}\right)^{\frac{}{}}$	34.04	241.74	781.26	1334.21	Out of range
Faulkenberry (2018) BIC-based BF	$BF_{01} = \sqrt{n^{df_1}\left(\frac{Fdf_1}{df_2}+1\right)^{-n}}$	179.94	982.97	3315.15	7687.04	708.78

Source: Result of simulation

As the number of observations per unit (m) grew from 5 through 100, the BIC-based Bayes factor (Faulkenberry, 2018) and the prior sensitive Bayes factor (Wang and Sun, 2013) indicated a steady rise in the weight of evidence in support of the null hypothesis of no between group variability. At no point did this two Bayes factors provided substantial evidence against the null hypothesis of no variability between the five factors/treatments. The BIC-based Bayes factor dropped suddenly at $m = 100$ whereas the Wang and Sun (2013) Bayes factor was not available at that point. From Table 4, at $m = 100$ the Wang and Sun (2013) Bayes factor reported an "Out of Range of the Gamma function".

CASE 2: Observation Per Units (m) Is Fixed While Factor/ Treatment Units (k) Are Increasing

The observations per treatment units are fixed at ($\boldsymbol{m = 10}$) units, whereas the number offactor/treatment units (k) will be increasing. Five (5) different factor sizes ($\boldsymbol{k = 5, 10, 20, 30\ and\ 35}$) were considered. Simulations were generated using random seeds to simplify replication. The methodologies discussed above are illustrated below using Case 2A corresponding to ($k = 5\ and\ m = 10$).

CASE 2A ($\boldsymbol{k = 5\ and\ m = 10}$):

Table 5: Simulated data for CASE 2A ($\boldsymbol{k = 5\ and\ m = 10}$)

		Observations per factor (m)									
Factors/Treatment Units (k)	A	-0.9	0.18	1.59	-1.13	-0.08	0.13	0.71	-0.24	1.98	-0.14
	B	0.42	0.98	-0.39	-1.04	1.78	-2.31	0.88	0.04	1.01	0.43
	C	2.09	-1.2	1.59	1.95	0	-2.45	0.48	-0.6	0.79	0.29
	D	0.74	0.32	1.08	-0.28	-0.78	-0.6	-1.73	-0.9	-0.56	-0.25
	E	-0.38	-1.96	-0.84	1.9	0.62	1.99	-0.31	-0.09	-0.18	-1.2

Source: Simulation Result

$$SST = \sum_{i=1}^{5}\sum_{j=1}^{10}(y_{ij} - \bar{y}_{..})^2 = 62.4978$$

$$SSB = m\sum_{i=1}^{5}(\bar{y}_{i.} - \bar{y}_{..})^2 = 2.2905 \qquad MSB = \frac{SSB}{k-1} = 0.5726$$

$$SSE = SST - SSB = 60.2073 \qquad MSE = \frac{SSE}{m(k-1)} = 1.3379$$

For the frequentist ANOVA, we seek to test the hypothesis,

$H_0 : \mu_1 = \mu_2 = ... = \mu_5$ against $H_1 : \mu_i \neq \mu_j$ for $i \neq j$ (3.4)

Table 6: Summary of One Way ANOVA table for CASE 2A $k = 5$ and $m = 10$

Source of Variation	Degree of Freedom (DF)	Sum of Squares (SS)	Mean Squares (MS)	F-Ratio	p-value (p)
Between Groups	4	2.2905	0.5726		
Within Groups (Error)	45	60.2073	1.3379	0.4280	0.7876
Total	49	62.4978			

Frequentist' Decision rule: Reject H_0 if $P_{value} < significance\ level\ (\alpha = 0.05)$

Frequentist' Conclusion:
Since $P_{value} = 0.7876 >$ the significance level $(\alpha) = 0.05$, we do not reject the null hypothesis stated in equation (2.4). This implies that the five treatment means are the same.

Wang and Sun (2013) Bayes Factor for Case 2A (k = 5 and m = 10)
We seek to test the hypothesis,

$H_0 : \sigma_\alpha^2 = 0$ against $H_1 : \sigma_\alpha^2 \neq 0$ (3.5)

The Wang and Sun (2013) Bayes factor for obtaining the weight of evidence in support of the null hypothesis for $(k = 5\ and\ m = 10)$ is computed as follows:

$$BF_{01} = \frac{\Gamma\left(\frac{mk-1}{2}\right)\Gamma(\alpha+1)}{\Gamma\left(\frac{k}{2}+\alpha+\frac{1}{2}\right)\Gamma\left(\frac{mk-k}{2}\right)}\left(\frac{SSE}{SST}\right)^{\frac{(mk-k-2)}{2+\alpha}}$$

$$= \frac{\Gamma\left(\frac{50-1}{2}\right)\Gamma\left(-\frac{1}{2}+1\right)}{\Gamma\left(\frac{5}{2}+\left(-\frac{1}{2}\right)+\frac{1}{2}\right)\Gamma\left(\frac{50-5}{2}\right)}\left(\frac{60.2073}{62.4978}\right)^{\frac{(50-5-2)}{2+\left(-\frac{1}{2}\right)}} = 241.74$$

Table 7: Decision Rule Table For Bayes Factor Interpretation

Bayes Factor (BF_{01})	Evidence For the null hypothesis (H_0)
$1-3$	Not worth more a mere mention
$3-10$	Substantial
$10-100$	Strong
> 100	Decisive

The Bayes factor $BF_{01} = 241.74$, signifies that the data has a decisive evidence in support of the null hypothesis of no variability between the five factors/treatments stated in equation (2.4). This can be seen in Table 7. Hence, its inverse $BF_{10} = 0.041$ indicates negligible evidence that the data could occur under the alternative hypothesis.

Faulkenberry (2018) BIC Based Bayes Factor for Case 2A $(k = 5\ and\ m = 10)$:
We seek to test the hypothesis,

$H_0 : \sigma_\alpha^2 = 0$ against $H_1 : \sigma_\alpha^2 \neq 0$ (3.6)

The Faulkenberry (2018) BIC-based Bayes factor for obtaining the weight of evidence in support of the null hypothesis for $(\boldsymbol{k = 5\ and\ m = 10})$ is computed as follows:

$$BF_{01} = \sqrt{n^{df_1}\left(\frac{Fdf_1}{df_2} + 1\right)^{-n}}$$

Where, n is the total sample size $(mk) = 50$. Hence,

$$BF_{01} = \sqrt{50^4\left(\frac{0.4280 \times 4}{45} + 1\right)^{-50}} = 982.97$$

The Bayes factor of $BF_{01} = 982.97$ indicates a decisive evidence in support of the null hypothesis of no variability between the five factors/treatments stated in equation (4.6). This can be seen in Table 7. Although, it's inverse $BF_{10} = 0.001$ indicates very negligible evidence that the data will occur under the alternative hypothesis.

Summary of Discussion for Case 2:

The methodologies illustrated in **case 2A** above were replicated for the four (4) other sub-cases under **case 2.** i.e

$\boldsymbol{k = 10\ and\ m = 10},\ \boldsymbol{k = 20\ and\ m = 10},\ \boldsymbol{k = 30\ and\ m = 10},\ \boldsymbol{k = 35\ and\ m = 10}$The results is summarized in **Table 8**:

Table 8: Summary of the Bayes factors proposed by Wang and Sun (2013) and Faulkenberry (2018) under Case 2

		$\boldsymbol{Case\ 1\ (m\ is\ fixed\ at\ 10\ while\ k\ is\ increasing)}$				
	Formular	**k=5**	**k=10**	**k=20**	**k=30**	**k=35**
Wang and Sun (2013) (BF)	$BF_{01} = \frac{\Gamma\left(\frac{mk-1}{2}\right)\Gamma(\alpha+1)}{\Gamma\left(\frac{k}{2}+\alpha+\frac{1}{2}\right)\Gamma\left(\frac{mk-}{2}\right.}$	241.74	1.40E5	1.05E10	1.58E13	Undefined
Faulkenberry (2018) BIC-based BF	$BF_{01} = \sqrt{n^{df_1}\left(\frac{Fdf_1}{df_2}+1\right)^{-n}}$	982.97	8.89E9	9.48E18	6.13E29	2.56E35

Source: Results of Simulation

With the number of observations per unit fixed at $m = 10$ and the number of factor/treatment units increasing from 5 through 35, the BIC-based and the prior sensitive Bayes factor increased in its evidence in support of the null hypothesis all through.

In all the cases studied, the Faulkenberry (2018) BIC-based Bayes factor proved to be more consistent in increasing weight of evidence in support of the hypothesis under test. Particularly, it has the same value at all sizes.

The Wang and Sun (2013) Bayes factor indicated "undefined" at $k = 35\ and\ m = 10$. This signifies that its inverse $BF_{10} = 0$. This means that there is entirely no evidence in support of the alternative hypothesis stated in equation (3.5) at that point.

Conclusion

In this work, we examined two Bayes factors (a prior sensitive and a BIC-based) under two cases (number of factors is fixed while observation per factor is increasing (i.e. random) and number of observations per factor is fixed while number of factors is increasing (i.e.

random)). In both scenarios, the two Bayes factors were consistent in increasing weight of evidence in support of the null hypothesis of "Zero Between Treatment Variability of the Factors/Treatments" been compared. As the sample size became large under the two cases, the prior sensitive Bayes factor became unavailable (impracticable). Specifically, under case 1, ($\boldsymbol{k = 5\ and\ m\ is\ random}$), the R software package used in obtaining the Bayes factor for testing the null hypothesis gave a response "out of range of Gamma function" at $m \geq 69$. Also under case 2, (□ = □□ □□□ □ □□ □□□□□□), the R software package used in obtaining the Bayes factor for testing the null hypothesis gave a response "out of range of Gamma function" at □ ≥ *35*.

In generalization, when the sample size becomes large, precisely at (□□ ≥ *350*), the prior sensitive Bayes factor proposed by (Wang and Sun 2013) has difficulty and in most cases becomes impracticable. A further examination reveals that its impracticality for large sample sizes lies in the Gamma function involved in its computation. Hence a need to develop an alternative to the Gamma functions.This confirms the assertion of other researchers (See, Faulkenberry, 2018) who stated that Bayes factors for more complex designs are quite nontrivial to compute, and such computation is an active area of research today (e.g., Nathoo and Masson, 2016).

References

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Faulkenberry, T.J. (2018). Approximating Bayes factors from minimal ANOVA summaries: An extension of the BIC method Department of Psychological Sciences Tarleton State University, USA arXiv:1710.02351v1 [stat.ME] 6 Oct 2017.

Gelman A. (2005). Analysis of variance: Why it is more important than ever. *Annals of Statistics,* 33, 1-31.

Janssen, D. P. (2011). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. Department of Psychology, University of Kent, Canterbury, Kent.UK.

Behav Res (2012). 44:232–247. DOI 10.3758/s13428-011-0145-1

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

Nathoo, F. S., & Masson, M. E. J (2016) Bayesian alternatives to null hypothesis significance testing for repeated-measures designs. *J Math Psychol, 7*2,144-157. doi:10.1016/j.jmp.2015.03.003.

Poirier, D. J. (2006). The Growth of Bayesian methods in statistics and economicssince 1970, *Bayesian Analysis*, 1, 969–980.

R Development Core Team. (2017), *R: A language and environment for statisticalcomputing*. Vienna, Austria: R Foundation for Statistical Computing,ISBN 3–900051–00–3.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*,111–163.

Steven, G., & Reinhard, F. (2013). Conjugate distributions in hierarchical Bayesian ANOVA For computational efficiency and assessments of both practical and statistical significance. University of Zurich, Switzerland.

Wang, M., & Sun, X. (2013). Bayes factor consistency for one-way random effects model. Department of Mathematical Sciences, Clemson University, USA. arXiv1302.6320v [stat.ME] 26 Feb. 2013.

Wetzels, R., Grasman, R., & Wagenmakers, E. (2012).Teachers corner; A default Bayesian hypothesis test for ANOVA designs. *104 © 2012 American Statistical Association DOI: 10.1080/00031305.2012.695956 The American Statistician,* 66(2).

Wike, E. L., & Church, J. D. (1976). Comments on Clark's the language-as-fixed-effect fallacy. *Journal of VerbalLearning and Verbal Behavior, 15,* 249-255.