The effect of outliers on the approximate adjusted fractional Bayes factor (AAFBF) in the ANOVA-context

Research Report

Marlyne Bosman

November 30, 2017

Supervisor :
Herbert Hoijtink

1 Introduction

Analysis of variance (ANOVA) is a statistical approach for comparing means that is used by many researchers. Just as any statistical approach, ANOVA is based on certain assumptions. When the data meet these assumptions, ANOVA can be validly applied. In reality, however, data often violates assumptions. For instance, sampling can result in an outlier, i.e. an observation that "deviates so much from other observations as to arouse suspicious that it was generated by a different mechanism" (Hawkins, 1980, p. 1).

Unfortunately, even a small proportion of outliers can seriously affect an ANOVA. Particularly, outliers cause an increased error variance, thereby leading to a reduced power of the statistical test (Wilcox, 2017). Additionally, outliers result in biased parameter estimates (Ruckstuhl, 2014; Wilcox, 2017). Therefore, if an ANOVA is applied to a dataset that contains outliers inference can be highly inaccurate.

One obvious way of handling the adverse effects of outliers is to remove them from the dataset prior to analysis. However, whether the removal of outliers is advisable depends on the source of the outliers. Ideally, one wants to keep outliers that are legitimate cases or for which the source is unknown in the data while at the same time minimizing their influence on estimation and hypothesis testing. One manner to achieve that objective is to use robust measures.

Robust measures are measures of central tendency and spread that are relatively unaffected by slight changes in a distribution (Wilcox, 2017, p.25).

Usually, non-robust measures, like the population mean, μ, and standard deviation, σ, are used to measure central tendency and spread of a distribution. However, in the presence of outliers μ and σ will result in inaccurate results. Conversely, robust measures will still give relatively accurate results (Ruckstuhl, 2014; Wilcox, 2017). A simple example of such a robust measure is the median. Unlike the mean, the value of the median is unaffected by outliers.

Robust measures are mostly discussed in the context of estimation and null hypothesis significance testing. Another approach for evaluating hypotheses is a Bayesian model selection approach (Klugkist, Laudy, & Hoijtink, 2005). This approach uses a Bayes factor (BF) to directly evaluate scientific expectations, stated as informative hypotheses. In the context of an ANOVA, an informative hypothesis can be used to state an expected ordering of means, for example,

$$H_1 : \mu_1 < \mu_2 < \mu_3, \tag{1}$$

where μ_j represent the mean of group $j = 1, 2, 3$. With the Bayes factor, the relative support in the data can be calculated for an informative hypothesis, H_i, compared with it's complement, H_c (van Rossum, van de Schoot, & Hoijtink, 2013), an unconstrained hypothesis, H_u, or another informative hypothesis, H_i'. For example, H_1, as stated in Equation 1, can be compared with another informative hypothesis,

$$H_2 : \mu_1 < \mu_2 = \mu_3. \tag{2}$$

Finding a BF_{12} of 5 indicates that the support in the data for hypothesis H_1 is five times larger than the support for hypothesis H_2.

Recently, Gu, Mulder, & Hoijtink (2017) developed the approximate adjusted fractional Bayes factor (AAFBF). With the AAFBF, informative hypotheses can be evaluated for virtually any statistical model. Additionally, the AAFBF is implemented in an easy-to-use software package called Bain. For the calculation, only the parameter estimates of the statistical model at hand and their covariance matrix are needed.

In the ANOVA context, the parameter estimates of interest are the group means. In a regular ANOVA, these are estimated by means of the Ordinary Least Squares (OLS) estimator. However, as previously stated, parameters that are estimated in a regular ANOVA can be seriously affected by outliers in the data. Hence, the expectation is that the AAFBF resulting from these estimates is also negatively affected by outliers. However, to our knowledge, this has never been formally investigated.

This paper aims to investigate to what extent the AAFBF based on the regular OLS estimates ($\text{AAFBF}_{\text{OLS}}$) is affected by outliers. Additionally, it aims to investigate to what extent replacing the OLS estimates as input for the AAFBF with robust estimates results in a decreased effect of outliers. For this purpose, a simulation study is used to show and compare the effect of outliers on the OLS estimator and an estimator that is relatively unaffected by outliers, i.e. a robust estimator. Subsequently, a simulation study will evaluate and compare the effect of outliers on the AAFB_{OLS} and the AAFBF based on a robust estimator.

This paper is organized as follows. Section 2 introduces a robust estimator suitable for the ANOVA context. Furthermore, it qualities are discussed and it is proposed to use the robust estimator for estimation of the parameters needed for BAIN. Next, Sections 3 and 4 describe the methods and results of the simulation study. Finally, in Section 5 the results, implications and limitations of the research are discussed.

2 Robust estimators

Wilcox (2017, Chapter 3) discusses the performance of various robust estimators. From this discussion, a robust estimator called the 20% trimmed mean emerges as a suitable estimator for the ANOVA context. The 20% trimmed mean is a robust measure of central tendency that deals with reducing the effect of outliers by removing 20% of a sample's distribution at both tails. It is calculated as follows,

$$\mu_t = \frac{1}{1-2\gamma}\int_{x_\gamma}^{x_{1-\gamma}} x dF(x), \tag{3}$$

where μ_t is the trimmed mean, $\gamma = 0.2$ is the amount of trimming and x_γ is the γth quantile. For the 20% trimmed mean, small-sample efficiency and accurate coverage probability are shown (Wilcox, 2017, Chapter 3).

The degree of resistance to outliers, i.e. robustness, of μ_t can be evaluated with it's influence function and breakdown point. As explained by Ruckstuhl (2014), the influence function of an estimator can be seen as a measure of local reliability: it measures the influence of the size of a single additional data point, e.g. y-value. Conversely, the breakdown-point can be seen as a measure of global reliability: it measures the maximum proportion of outliers for which an estimator still returns reliable estimates.

In Figure 1, an illustration of the empirical influence functions of μ and μ_t (given in Wilcox, 2017) is shown for a sample of size 65 from the standard normal distribution. The empirical influence function, IF(y) quantifies the effect of adding an y-value to the dataset with an arbitrary large value has on the value of the population mean estimate. As can be seen in Figure 1, the value of the sample mean increases without bounds if an increasingly outlying y-value is added to the data. Meanwhile, μ_t has a bounded influence function: an additional y-values that is too extreme is trimmed and cannot influence the estimated parameter.

Furthermore, the breakdown point of μ_t is 0.2, that is, if less than 20% of the values of a data set are outliers, μ_t will still return a relatively reliable estimate of the central tendency of the data (Wilcox, 2017, p.39). By way of contrast: the breakdown point of the mean is 0.0, i.e. one outlier can cause the mean to go to infinity. The highest possible breakdown point is 0.5, which is for example achieved by the median. While μ_t does not have the highest possible breakdown point, accurate coverage probability is considered to be a more important quality of a robust estimator. Based on the described performance of μ_t, this paper proposes to use μ_t as input for `Bain` for the calculation of Bayes factors evaluating informative hypotheses, resulting in AAFBF_{μ_t}.

3 Methods

General

Data will be simulated in R (R Core Team, 2016) from the ANOVA model:

$$y_i = \sum_{j=1}^{3} \mu_j D_{ij} + \epsilon_i \tag{4}$$

where y_i is the observation on the dependent variable for person i ($i = 1, ..., N$), where N denotes the sample size, μ_j denotes the mean of group

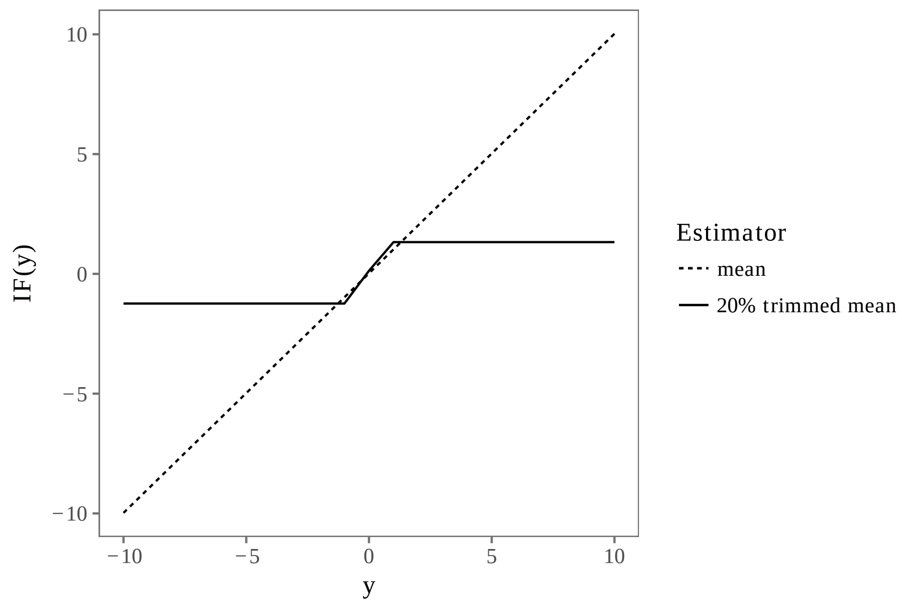

Figure 1: The empirical influence function of the mean and the 20% trimmed mean for a random sample ($n = 65$) drawn from the standard normal distribution. Illustrated is the effect adding an outlier varying between $-10 < y < 10$ on the value of the parameter estimate.

j ($j = 1, 2, 3$), $D_{ij} = 1$ if person i is in group j and 0 otherwise and ϵ_i denotes the residual for person i, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ where σ^2 denotes the residual variance. For simulated data, the following informative hypotheses will be evaluated by means of the AAFBF:

$$H_1 : \mu_1 = \mu_2 = \mu_3 \tag{5}$$

$$H_2 : \mu_1 < \mu_2 < \mu_3 \tag{6}$$

$$H_u : \mu_1, \mu_2, \mu_3 \tag{7}$$

where H_u denotes the unconstrained hypothesis, i.e. a hypothesis with no constraints on the means.

Two populations are considered, in the first H_1 is true and in the second H_2 is true. For the first population, the group means are chosen to be $\mu_1 = \mu_2 = \mu_3 = 0.0$ with residual variance $\sigma^2 = 1.0$. For the second population, the group means are chosen to be $\mu_1 = 0.0, \mu_2 = 0.5, \mu_3 = 1.0$ with residual variance $\sigma^2 = 1.0$, corresponding to a medium effect size between two groups (Cohen, 1988).

From both populations $1,000$ data sets are sampled with sample size per group $n_j = 65$. The sample size is chosen such to have enough power (80%) to detect a difference of medium effect size with a regular ANOVA. For each data set, the OLS estimates of the population means and their covariance matrix are calculated by means of the R base function for fitting a linear model `lm()`. Additionally, for each data set, the 20 % trimmed mean estimate is calculated by means of the R base function `mean()` for which the trimming argument is set equal to 0.2. The standard error for the 20% trimmed mean is calculated by means of the function `trimse` from the WRS2 package (Mair, Schoenbrodt, & Wilcox, 2017). From the standard error of the 20% trimmed mean, the variance is calculated by taking the standard error to the power 2.

Subsequently, resulting estimates and their covariance matrix are fed into the `Bain` function from the BAIN package for the calculation of the Bayes factors evaluating the informative hypotheses stated in Equations 5-7. For each dataset, this will result in the following three Bayes factors: BF_{1u}, BF_{2u} and BF_{12} evaluate the relative support in the data for H_1 compared to H_u, H_2 compared to H_u and H_1 compared to H_2 respectively.

Performance evaluation

From the resulting distribution of OLS estimates and 20% trimmed mean estimates, absolute bias, δ, is calculated following

$$\delta = |(\sum_{i=1}^{R} \frac{\hat{\mu}}{R}) - \mu| \tag{8}$$

in which $R = 1,000$ is the number of simulated data sets, $\hat{\mu}$ is either the OLS or 20% trimmed mean estimate and μ is the population mean. Additionally, the coverage probability of the 95% confidence interval (CI) of each estimate is calculated by counting how often the population mean is in the interval, whereby the limits of the 95% CI are calculated following

$$\hat{\mu} \ \pm \ t_{0.975} \ SE \tag{9}$$

in which $t_{0.975}$ is from a Student's t distribution with $n_j - 1$ degrees of freedom for the OLS estimate and $n_j - 2\gamma n_j - 1$ degrees of freedom for the 20% trimmed mean estimate (Wilcox, 2017, Chapter 4) and SE is the standard error of the estimate. Finally, from the resulting distribution of Bayes factors, the mean will be calculated.

Outliers

The in the last section described performance evaluations will be repeated after values in the data sets are replaced by outliers. As it is beyond the scope of this research to investigate all potential situations concerning the generation of outlying values, one specific, realistic situation is chosen as the focus of this research. This paper focusses on the situation where outliers occur only in one group (randomly chosen to be group 3) due to motivated under-reporting (Osborne & Overbay, 2004). Since the breakdown point of μ_t is known to be 0.2, up to 20% outliers are considered. The size of the outliers is based on the robust MAD-Median rule for detecting outliers (Wilcox, 2017, p.101): data values are replaced by values that would be detected as an outlier by the MAD-Median rule.

4 Results

Figure 2 shows the relationship between the number of outliers and size of δ for both the OLS and 20% trimmed mean estimate of $\mu_3 = 0.0$ in the first population. As can be seen in Figure 2, an increase in the number of outliers coincides with an increase in δ for both estimators. However, the increase in bias is considerably smaller for the 20% trimmed mean estimator.

Subsequently, Figure 3 shows the relationship between the number of outliers and the coverage probability of the 95% CI's for both the OLS and 20% trimmed mean estimates of $\mu_1 = \mu_2 = \mu_3 = 0.0$ in the first population. As can be seen in Figure 3, an increase in the number of outliers coincides with a decrease in the coverage probability for group 3 for both estimators. For the OLS estimate, this drop in coverage probability is enormous and already occurs with a few outliers. For the 20% trimmed mean estimate, coverage probability remains relatively reasonable. Additionally, Figure 3 shows an increased coverage probability as a function of number of outliers

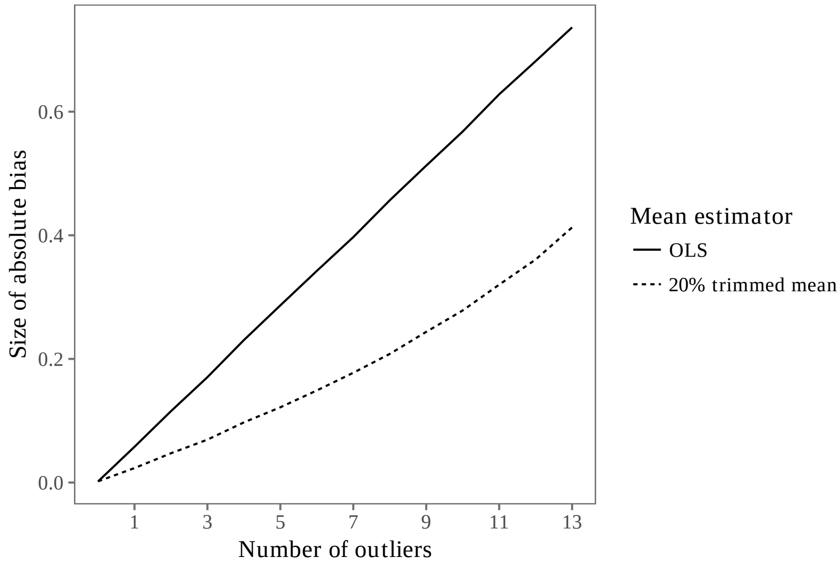

Figure 2: The relationship between the number of data values replaced by outliers and absolute bias of the OLS and 20% trimmed mean estimate of the population mean for group 3 in the first population ($\mu_3 = 0.0$).

for group 1 and 2 for the OLS estimator. This is the result of an increased error variance as a consequence of outliers in combination with the equal variances assumption from the regular ANOVA, resulting in a 95% CI that is too wide.

Figure 4 shows the relationship between the number of outliers and the size of the Bayes factors for the first population. As can be seen in Figure 4 the support in the data for the true hypothesis (H_1) compared to H_u decreases as a function of the number of outliers for both estimators. For the AAFBF$_{\text{OLS}}$ the decrease is stronger than for the AAFBF$_{\mu_t}$. However, up till about 9 10 (14 15%) outliers, BF_1u still indicates quite some support in the data for H_1 compared to H_u for both AAFBF. Additionally, while the support in the data for H_1 compared to H_u decreases, the support for H_1 compared to H_2 increases. Apparently, with more outliers, the truth is even further away from H_2, even though H_1 also does not captures it.

Finally, Figure 5 shows the relationship between the number of outliers and the size of the Bayes factors for the second popualation. As can be seen in Figure 5 the support in the data for the true hypothesis (H_2) compared to H_u decreases as a function of the number of outliers for both estimators. Again, for the AAFBF$_{\text{OLS}}$ the decrease is stronger than for the AAFBF$_{\mu_t}$. However, also the AAFBF$_{\text{OLS}}$ is able to identify H_2 as the best hypothesis

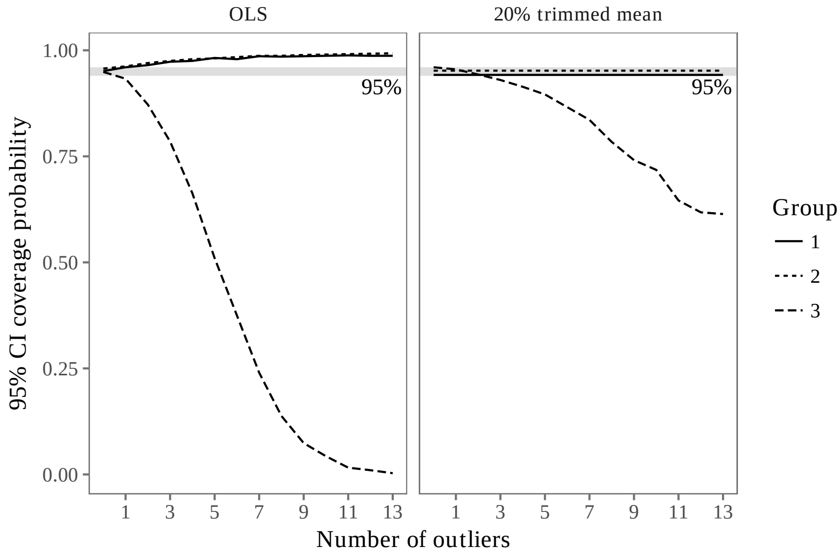

Figure 3: The relationship between the number of data values in group 3 replaced by outliers and coverage probability of the 95% CI around the OLS and 20% trimmed mean estimate of the population mean for group 1, 2 and 3 ($n_j = 65$). Note that outliers are only in group 3.

up till about 9 (14%) outliers. With 10 or more outliers, H_1 is considered more probable compared to H_2 by the $\text{AAFBF}_{\text{OLS}}$.

5 Discussion

From the simulation results, one could conclude that the $\text{AAFBF}_{\text{OLS}}$ seems quite robust. For both situations, up to about 14% outliers, $\text{AAFBF}_{\text{OLS}}$ selects the true hypothesis as the best hypothesis. However, since only one situation of outliers is investigated in this research and the simulation results also show that the size of the $\text{AAFBF}_{\text{OLS}}$ is strongly influenced by the number of outliers, further research is necessary to fully understand the effect of outliers on the AAFBF.

Further inspection of the results showed with the OLS estimates the ordering of means did not change up to about 14% outliers. This suggests that the effect of outliers might be reinforced for a smaller effect size in the population, which would cause a change in the estimated ordering of means for a smaller number of outliers. Additionally, the direction of the outliers in combination with in which group outliers are present influence estimated ordering of means. Furthermore, for other true orderings of means, the effect of outliers might also be different than the effect measured in the present

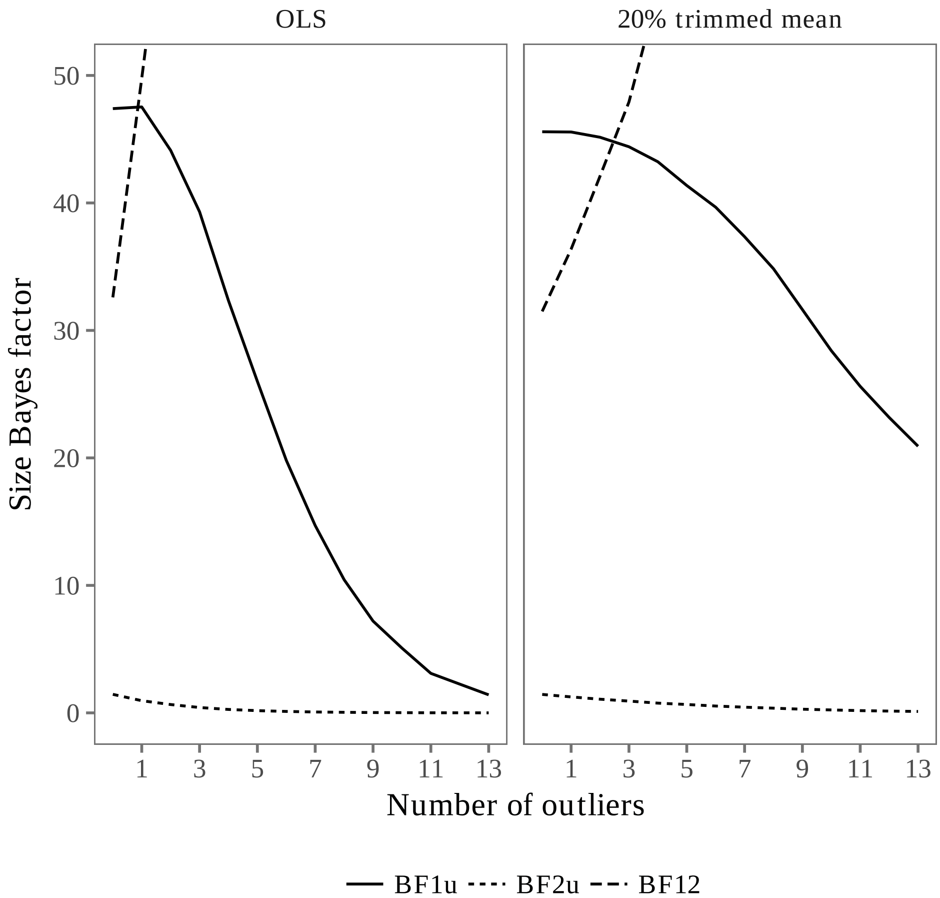

Figure 4: The relationship between the number of data values in group 3 replaced by outliers and the size of the Bayes factors evaluating the relative support in the data for the informative hypotheses stated in Equations 5-7, with the OLS or 20% trimmed mean estimates as input. The underlying truth is captured in H_1, stated in Equation 5. Outliers in group 3 are on the left tail of the distribution. Note that for the sake of visibility, the y-axis has been cut at BF = 50.

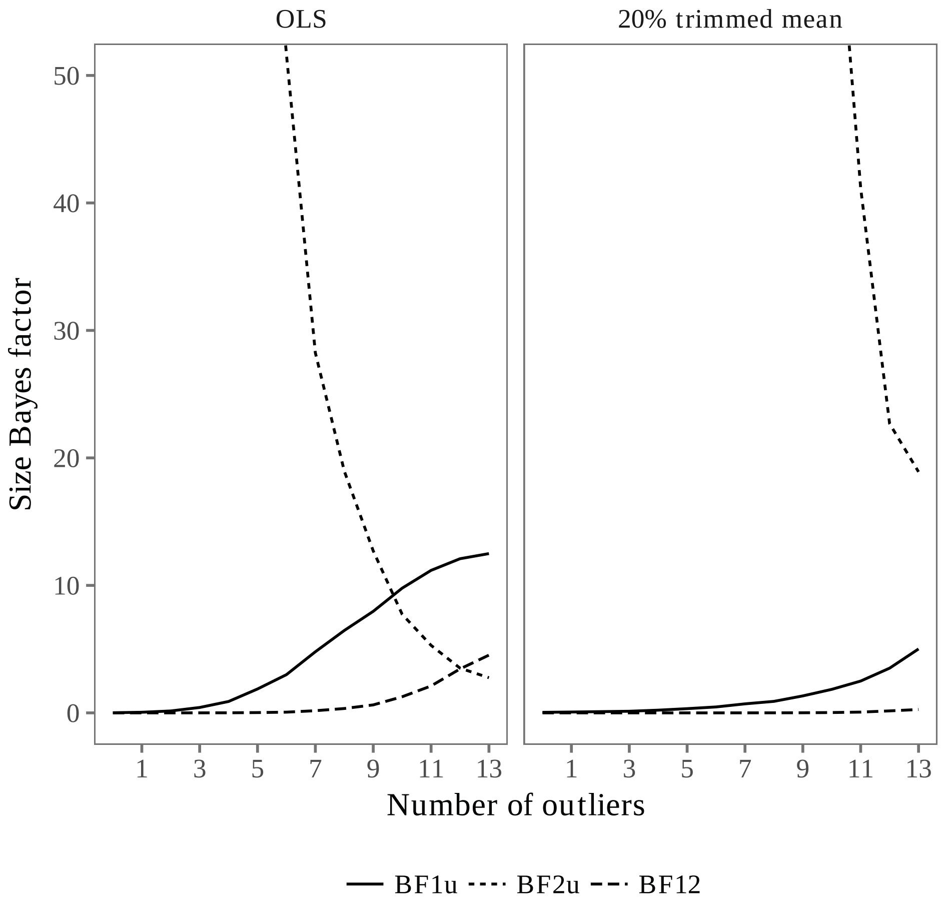

Figure 5: The relationship between the number of data values in group 3 replaced by outliers and the size of the Bayes factors evaluating the relative support in the data for the informative hypotheses stated in Equations 5-7, with the OLS or 20% trimmed mean estimates as input. The underlying truth is captured in H_2, stated in Equation 6. Outliers in group 3 are on the left tail of the distribution. Note that for the sake of visibility, the y-axis has been cut at BF = 50.

research. We expect the AAFBF_{μ_t} to be more robust against these influences than the $\text{AAFBF}_{\text{OLS}}$. If this is indeed the case could be investigated by future research.

Conclusion

In conclusion, the AAFBF based on the OLS estimates seems quite robust in the situations simulated in this research. However, the AAFBF based on the 20% trimmed mean estimates performs even better. In addition, some sample and population characteristics, like direction of outliers or effect size, might reinforce the adverse effect of outliers on the behaviour of the AAFBF based on the OLS estimates. Therefore, we would advise researchers that worry about the potential impact of outliers in their dataset to use a robust estimator, like the 20% trimmed mean, for the calculation of the AAFBF. Future research should provide us with more clarity with regard the effect of the interaction between sample and population characteristics and outliers on the behaviour of Bayes factors.

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gu, X., Mulder, J., & Hoijtink, H. (2017). Approximated Adjusted Fractional Bayes Factors: A General Method for Testing Informative Hypotheses. *British Journal of Mathematical and Statistical Psychology*. doi: 10.1111/bmsp.12110

Hawkins, D. M. (1980). *Identification of outliers* (11th ed.). London: Chapmann and Hall.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality Constrained Analysis of Variance: A Bayesian Approach. *Psychological Methods*, *10*(4), 477–493. doi: 10.1037/1082-989x.10.4.477

Mair, P., Schoenbrodt, F., & Wilcox, R. (2017). WRS2: Wilcox Robust Estimation and Testing [Computer software manual]. (0.9-2) doi: 10.1002/9781118445112.stat06356.pub2

Osborne, J. W., & Overbay, A. (2004). The Power of Outliers (and Why Researchers Should Always Check for Them). *Practical assessment, research & evaluation*, *9*(6), 1–12. Retrieved from `http://pareonline.net/htm/v9n6.htm`

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ruckstuhl, A. (2014). Robust Fitting of Parametric Models Based on M-Estimation. Retrieved from https://stat.ethz.ch/wbl/wbl4/WBL4_robstat14E.pdf

van Rossum, M., van de Schoot, R., & Hoijtink, H. (2013). "Is the Hypothesis Correct" or "Is it Not". *Methodology*, *9*(1), 13–22. doi: 10.1027/1614-2241/a000050

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th ed.). New York: Academic press.