Bayesian ANOVA in the presence of non-normality or outliers

Master's Thesis

Marlyne Bosman

Master's Programme:
Methodology and Statistics for the Behavioural,
Biomedical and Social Sciences, Utrecht University

Supervisor :
Herbert Hoijtink

January 12, 2018

1 Introduction

Analysis of variance (ANOVA) is a statistical approach for comparing means that is used by many researchers. ANOVA can be validly applied when the data meets certain assumptions. In reality, however, data often violates assumptions. For instance, distributions are rarely normal. Specifically, in their research of real samples both Micceri (1989) and Cain et al. (2016) frequently encountered distributions that were asymmetric (skewed) or had more observations in the tails (heavy-tailed). Moreover, random sampling can result in an outlier, i.e. an observation that "deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism" (Hawkins, 1980, p. 1).

Even small departures from normality or a small proportion of outliers can seriously affect an ANOVA. Particularly, heavy-tailed distributions and outliers cause increased error variance, thereby leading to reduced power of the statistical test (Wilcox, 2017, p. 2). Furthermore, skewed distributions cause inaccurate confidence intervals around the means (Wilcox, 2017, p. 5). Additionally, outliers result in biased parameter estimates (Ruckstuhl 2014; Wilcox 2017, p. 7). Hence, if an ANOVA is applied to a dataset that contains non-normality or outliers, inference can be highly inaccurate.

Since non-normality and outliers lead to inaccurate statistical inference, researchers should pay attention to how they want to handle potential non-normality and outliers in their data. The objective when handling the adverse effects of outliers seems quite clear. Outliers are either legitimate cases sampled from the tail of the distribution or illegitimate cases (Osborne & Overbay, 2004). Either way, both type of outliers distract from the truth. Hence, their influence on estimation and hypothesis testing should be minimized. The objective when handling the adverse effects of non-normality is less clear because the truth is less obvious. For instance, which statistic tells us the truth when a distribution is skewed? However, in the context of comparing groups, a method that more accurately describes the bulk of the data and does not loose so much power when the data are heavy-tailed seems preferable. One manner to achieve these objectives is to use robust statistical inference.

Robust statistics are measures of central tendency and spread that are unaffected by slight changes in a distribution (Wilcox, 2017, p. 25). Usually, non-robust statistics, like the mean, μ, and standard deviation, σ, are used to measure central tendency and spread of a distribution. Indeed, these statistics are also the basis of an ANOVA. However, in the presence of non-normality or outliers μ and σ will be inaccurately estimated. Conversely, robust statistics will still give relatively accurate results (Ruckstuhl 2014; Wilcox 2017, pp. 25-31). A simple example of such a robust statistic is the median. Unlike the mean, the value of the median is unaffected by a single outlier.

Robust statistics are mostly discussed in the context of estimation and null hypothesis significance testing, but not in the context of the Bayesian model selection approach. The Bayesian model selection approach (Klugkist, Laudy, & Hoijtink, 2005) 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 represents 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 estimates and covariance matrix of the parameters of the statistical model at hand 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 estimates 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 ($\text{AAFBF}_{\text{ROB}}$) results in a decreased effect of outliers. The paper is organized as follows. Section **??** introduces a robust estimator suitable for the ANOVA context. A simulation study is set up to show and compare the effect of outliers on the OLS estimator and the robust estimator and equally for the AAFB_{OLS} and AAFB_{ROB}. Sections **??** and **??** describe it's set-up and results. Finally, in Section **??** the results, implications and limitations of the research are discussed.

References

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2016). Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation. *Behavior Research Methods*, *49*(5), 1716–1735. doi: 10.3758/s13428-016-0814-1

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

Micceri, T. (1989). The Unicorn, the Normal Curve, and Other Improbable Creatures. *Psychological Bulletin*, *105*(1), 156–166. doi: 10.1037/0033-2909.105.1.156

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`

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.