8-9-2019

The (Mis)use of One-Way ANOVA Testing in Sociology Research Papers and Their Implications

Nancy Yang
Portland State University

The (Mis)use of One-Way ANOVA Testing in Sociology Research Papers and Their Implications

by

Nancy Yang

An undergraduate honors thesis submitted in partial fulfillment of the

requirements for the degree of

Bachelor of Arts

in

University Honors

and

Mathematics

Thesis Adviser

Dr. Nadeeshani Jayasena

Portland State University

2019

Table of Contents

1. Introduction

Many research papers being produced today utilize p-values and hypothesis testing (John Ioannidis 2019). Null hypothesis statistical testing (NHST) is very useful for its seemingly easy to interpret results, however, with the increase in the use of NHST and the explicit inclusion of p-values in research papers, there have been a number of articles, blogs, and other written works critiquing the use and validity of these test results including works by Ioannidis (2019), Nahm (2017), Glantz (1980), Engman (2011), Dar et al (1994), McLean & Ernest (1998), and Nickerson (2000). These works, and many others specifically speak on the inclusion and mishandling of p-values in research.

Statistical testing requires certain assumptions to ensure the most accurate results. For the Analysis of Variance (ANOVA) test, a type of Null Hypothesis test, the assumptions are that the provided data are normally distributed, randomly selected, independent, and different groups have the same variance. ANOVA is robust to normality, meaning that it can still produce accurate results despite violating the assumption of normality (Laerd Statistics Accessed 2019). Violating this assumption can also have minimal effect on the Type I error rate (Laerd Statistics Accessed 2019).

"Statistically significant" is a phrase used by researchers to determine the status of their hypothesis. A p-value that is smaller than the significance level (often 0.05) is considered statistically significant. Some researchers may report this "significance" with no basis on what that could mean in relation to their question and/or research (Engman 2011). This may be related to the nature of journal publication requirements in which statistically significant results need to be reported (Reinhart 2015; Nickerson 2011).

The research around the use of NHST in research papers is dominated by statisticians and biomedical researchers (Yoccoz 1991), however, it is acknowledged that the use of NHST in research is increasing over time in the social sciences (Engman 2011). This paper aims to understand the misuse of ANOVA testing in the social sciences, specifically the area of sociology and related disciplines. To reach this understanding, this paper will provide the guidelines and methods of our research, initial findings, discussion of incorrect interpretations, and finally, potential solutions to rectify these implications.

2. Methodology

This research will focus on the misuse of one-way ANOVA in sociology research papers. Other forms of ANOVA are more complex with varying degrees of complexity. One-way ANOVA is among the simpler forms which can act as a baseline for future research in other forms of ANOVA. This study considers whether the papers mentioned and/or checked the validity of one-way ANOVA assumptions, however, the main focus of evaluation is if and how the results and p-values are interpreted in relation to their initial questions and/or hypotheses.

Research articles are gathered using the SocINDEX database for peer reviewed articles and using keywords "one-way ANOVA". The scope of this study is within the years 2014-2019. Searches were conducted for each year and were compiled based on availability and their misinterpretation or omission of interpretation of ANOVA and its results (see Bibliography).

For the ANOVA test, the null hypothesis is that the means of the groups being studied are the same. Alternatively, the alternative hypothesis for ANOVA is that at least one of the group means being studied is different. For the NHST, the correct decision rule based on the p-values is as follows. When the p-value is larger than the significance level, we fail to reject the null hypothesis and cannot conclude that the data supports the alternative hypothesis. Likewise, if the p-value is smaller than or equal to the significance level, we reject the null hypothesis and conclude that the data supports the alternative. As a result, misinterpretations of results can come in similar forms of the following: the null is supported by the data or the null is true when the p-value is higher than the significance level.

3. Initial Findings

Overall, there is a general increase of search results produced from each year, supporting Engman's claims about the increase of the use of hypothesis testing in the social sciences (2011). There was a total of 8 results from the 2014 search for sociology papers using one-way ANOVA. The highest number of results found were in 2016 with 123 articles. In the 2017 search, there were 64 articles found, and an increase in articles were

found for each year following (excluding the current year). These results show that the use of one-way ANOVA has increased over the years but does not comment on the number of research papers containing misinterpreted results for each year. Overall, 13 papers were gathered based on their interpretations, however the quantity for each year sheds little light on the magnitude of errors being made per year.

The assumptions for ANOVA were generally unacknowledged within the collected sample of research papers. Papers that acknowledged the checking of assumptions were those in which violations to the assumptions were made. Wilson (2018) employed the Levene's test to check the equality of variances and found their groups contained unequal variances. Accordingly, Wilson (2018) conducted the Welch's test for unequal variances. This procedure was consistent among other papers noting the assumptions of one-way ANOVA. These papers contained no misinterpretations or errors, and thus were not collected.

Findings with no statistical differences proved to be a difficult point of interpretation in the observed articles. The most common misinterpretation found was a statement regarding a lack of significant differences being found. One interpretation was the confirmations of or proof for the given null hypothesis. For example, one article stated, "The t-test and ANOVA proved that the teachers do not vary in the level of their stereotypic belief with their demographic background variables" (Tafa 2016). Byrne et al (2014) and Hreish et al (2017) also contained conclusions of a true or confirmed null hypothesis.

Similar to proving the null hypothesis, papers contained statements of "no differences found" and "groups were similar/agreed" in relation to finding no statistically significant differences. These papers found their data to be not statistically significant in the results section and later referred to these findings as groups not differing from one another (Smith-Genthôs et al 2016; Lopes & Rusi 2015; Allen 2015; Gilmore et al 2016; Skinner & McHale 2016; Rawlings & Blackmer 2018; Kenny et al 2019).

One article confirmed their null hypothesis and cited similar results to the study it was largely based on (Hreish et al 2017). Upon analysis, it was found that this interpretation was not a result of interpretations in the cited source.

There was an attempt to gather articles solely with misinterpreted results, however it became evident that in-depth interpretations were omitted or simply forgotten.

Khosravan et al (2014) and Stanton et al (2017) noted the use of one-way ANOVA in their research. Their results were considered to not be statistically significant. Further comment or explanation about the relation of these results to the question were omitted.

4. Discussion

Since assumptions were generally not mentioned, it is difficult to glean any further understanding about the procedural misuse of one-way ANOVA testing in sociology research papers.

When finding no significant differences in ANOVA testing, the resulting decision rule is a failure to reject the null hypothesis. This indicates that there is not enough evidence to support the alternative hypothesis, or that there is no evidence that the group means are different. This does not prove that group means are the same. Therefore, the claims of proved or confirmed null hypotheses are not true. Although the data produce statistically significant results, we cannot claim that hypothesis is confirmed because it simply tells us if our data support the alternative claim (Nahm 2017). Additionally, when conducting a NHST, it is assumed that the null hypothesis is true. The purpose of conducting an ANOVA is to determine whether the data show evidence opposed to this assumption about the null hypothesis. Therefore, claims of a true null hypothesis or claims of no differences among groups are incorrect.

Statistical testing can be useful in helping us understand our data, but this does not work when they stand alone. These numbers and results need interpretation so the reader can better understand the purpose of these tests in relation to their findings. More explicit interpretations may also contribute to further study. An attempt was made to analyze how incorrect interpretations could lead to complications in future research. The main focus for this was to determine proceeding articles citing results (or similar results) from papers with misinterpreted p-values. There was not sufficient evidence to move forward with this analysis. A possible explanation for this is the incorrect interpretation following no significant differences in research papers. When claiming that there were no differences, it can suggest that no further interesting data can be found. On the contrary, it should suggest

that there is not enough information yet or methods of the given research can be improved. As a result, these results should lead to more analyses from different perspectives.

In articles that do not address statistically significant results further than their significance, it is difficult to understand the author's stance on their findings. Although the decision rule is met, the interpretations are lacking and thus does not add to the overall argument for each paper. This causes researchers to use statistical testing to no benefit. This omission of interpretation can potentially be the result of mathematical anxiety noted in Deckard's paper on the statistical education of sociology majors (2017). Deckard notes that the reasoning for some students in choosing social scientific fields is due to mathematical anxiety. It is possible that this mathematical anxiety negatively affects a researcher or student's ability to completely evaluate statistical results, however further study needs to be done to support this claim.

5. Potential Solutions

The common interpretation of "no statistical differences" in the one-way ANOVA is easily amended by concluding that there is no sufficient evidence to support claims against the null hypothesis, or there is no sufficient evidence to show differences among group means. This statement can be further improved by suggesting different ways of looking at the initial question at hand to spark interest in future research.

In cases where statistical significance is mentioned, but not evaluated, it may be important to understand the reasoning behind the omission. If a researcher finds that the findings are not beneficial to their study, it may be best to leave this information out or add commentary about the lack of benefit that the findings add to the resulting paper.

Lastly, it is important to incorporate statistical assessment by a qualified statistician in the review process for each journal accepting research containing statistical data. A lack in knowledge about the importance and effect of statistical data in the review process may lead to further misinterpretations in the future. For example, requiring researchers to include statistically significant results might not add as much substance as is imagined (Reinhart 2015).

6. Limitations

Our results are heavily based on limitations. In addition to a lack in time and resources, our search results were limited by articles produced from a single database, and reviewed articles were limited by availability through Portland State University credentials. It is acknowledged that, as a result of these limitations, claims from this research may be skewed or biased. There were only a small number of articles that were able to be gathered. Thus, the articles that were gathered for the purpose of evaluation in this study may not properly represent all sociology related articles using one-way ANOVA. Further study with different databases, specific journals, etc. are encouraged. Furthermore, a more in-depth analysis about the use and effects of unchecked or violated assumptions of one-way ANOVA in research should be considered.

7. Conclusion

Although this study did not yield expected results, it provides evidence that more research and evaluation of current approaches to statistical testing need to be made. Deckard (2017) noted that mathematical anxiety is an issue that some social scientists may face that lead them to the social sciences. This mathematical anxiety could be a reason for the omission of interpreted results of the ANOVA, however, further research needs to be done. Journal guidelines also pose a question for how valid results are in published articles; do these statements of statistical significance really add value to their study? (Reinhart 2015; McLean & Ernest 1998). Further research should be done focusing on specific journals.

Bibliography

References

1. Dar, Reuven, Ronald C. Serlin, and Haim Omer. "Misuse of Statistical Tests in Three Decades of Psychotherapy Research." *Journal of Consulting and Clinical Psychology*62, no. 1 (February 1994): 75-82. doi:10.1037//0022-006x.62.1.75.
2. Deckard, Natalie. "Statistics Education for Undergraduate Sociology Majors: Survey Findings across Institutions." *Numeracy*10, no. 2 (2017). doi:10.5038/1936-4660.10.2.8.
3. Engman, Athena. "Is There Life after P." *Quality & Quantity*47, no. 1 (June 01, 2011): 257-70. doi:10.1007/s11135-011-9516-z.
4. Glantz, S. A. "Biostatistics: How to Detect, Correct and Prevent Errors in the Medical Literature." *Circulation* 61, no. 1 (January 1980): 1-7. doi:10.1161/01.cir.61.1.1.
5. Ioannidis, John PA., What Have We (Not) Learnt from Millions of Scientific Papers with *P*-Values?, 2019, The American Statistician, 73:1,20-25.
6. McLean, James E., and James M. Ernest. "The Role of Statistical Significance Testing In Educational Research." *Research in the Schools*5, no. 2 (1998): 15-22.
7. Nahm, Francis Sahngun. "What the P Values Really Tell Us." *The Korean Journal of Pain* 30, no. 4 (October 2017): 241-42. Accessed June 2019. doi:10.3344/kjp.2017.30.4.241.
8. Nickerson, Raymond S. "Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy." *Psychological Methods*5, no. 2 (2000): 241-301. doi:10.1037/1082-989x.5.2.241.
9. "One-way ANOVA (cont...)." One-way ANOVA - Violations to the Assumptions of This Test and How to Report the Results | Laerd Statistics. Accessed June 04, 2019. https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-3.php.
10. Reinhart, Alex. *Statistics Done Wrong: The Woefully Complete Guide*. Daly City, CA: No Starch Press, US, 2015.
11. Yoccoz, Nigel G. "Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology." *Bulletin of the Ecological Society of America* 72, no. 2 (June 1991): 106-11. https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/20167258.

Articles Reviewed

1. Allen, Kathleen P. ""We Don't Have Bullying, But We Have Drama": Understandings of Bullying and Related Constructs Within the Social Milieu of a U.S. High School." *Journal of Human Behavior in the Social Environment*25, no. 3 (March 04, 2015): 159-81. doi:10.1080/10911359.2014.893857.
2. Wilson, Bincy. "A Women-Centric Approach: Key to Satisfaction With Agency Services." *Affilia*34, no. 2 (November 19, 2018): 219-36. doi:10.1177/0886109918806272.
3. Byrne, Rebecca, Anthea Magarey, and Lynne Daniels. "Food and Beverage Intake in Australian Children Aged 12-16 Months Participating in the NOURISH and SAIDI Studies." *Australian and New Zealand Journal of Public Health*38, no. 4 (August 5, 2014): 326-31. doi:10.1111/1753-6405.12249.
4. Gilmore, Amanda K., Hollie F. Granato, Sarah M. Wilson, and William H. George. "Sexual Assault and Heavy Episodic Drinking Among Women of Asian/Pacific Islander Ancestry and Women of European Ancestry." *Psychology of Women Quarterly*40, no. 3 (2016): 441-50. doi:10.1177/0361684316648312.

5. Hreish, Khalid, Mohammad Okkeh, Anan J. Fareed, and David S. Byers. "Attitudes among Young Adults in Palestine about Peers with Substance Use Problems: Challenges and Opportunities for Community Intervention Design." *International Social Work*62, no. 2 (December 19, 2017): 726-40. doi:10.1177/0020872817743562.

6. Kenny, Maureen C., Claire Helpingstine, Haiying Long, Lorena Perez, and Maria Clara Harrington. "Increasing Child Serving Professionals' Awareness and Understanding of the Commercial Sexual Exploitation of Children." *Journal of Child Sexual Abuse*28, no. 4 (February 14, 2019): 417-34. doi:10.1080/10538712.2018.1563264.

7. Khosravan, Shahla. "Epidemiology of Loneliness in Elderly Women." *Journal of Research & Health*4, no. 4 (2014): 871-77.

8. Lopes, Barbara, and Rusi Jaspal. "Paranoia Predicts Out-group Prejudice: Preliminary Experimental Data." *Mental Health, Religion & Culture*18, no. 5 (2015): 380-95. doi:10.1080/13674676.2015.1065475.

9. Rawlings, Mary A., and Emily R. Blackmer. "Assessing Engagement Skills in Public Child Welfare Using OSCE: A Pilot Study." *Journal of Public Child Welfare*13, no. 4 (August 23, 2018): 441-61. doi:10.1080/15548732.2018.1509760.

10. Skinner, Olivenne D., and Susan M. McHale. "Parent–Adolescent Conflict in African American Families." *Journal of Youth and Adolescence*45, no. 10 (October 2016): 2080-093. doi:10.1007/s10964-016-0514-2.

11. Smith-Genthôs, K. Rachelle, Erin M. Logue, Blakely E. Low, and Susan S. Hendrick. "The Forgotten Ones: Siblings of Substance Abusers." *Journal of Loss and Trauma*22, no. 2 (2017): 120-34. doi:10.1080/15325024.2016.1202005.

12. Stanton, Alexis G., Morgan C. Jerald, L. Monique Ward, and Lanice R. Avery. "Social Media Contributions to Strong Black Woman Ideal Endorsement and Black Women's Mental Health." *Psychology of Women Quarterly*41, no. 4 (December 7, 2017): 465-78. doi:10.1177/0361684317732330.

13. Tafa, Aseffa. "Teachers' Notion of the Sources of Stormy and Stressful Behaviors of Adolescents: Ethiopian Upper Primary School Teachers in Focu." *Indian Journal of Community Psychology*12, no. 2 (September 1, 2016).