

Robust Within Groups Anova: Dealing With Missing Values

Jinxia Ma, Rand R. Wilcox*

Dept. of Psychology, University of Southern California, Los Angeles, CA 90089-1061, United States

*Corresponding Author: rwilcox@usc.edu

Copyright ©2013 Horizon Research Publishing All rights reserved.

Abstract The paper considers the problem of testing the hypothesis that $J \geq 2$ dependent groups have equal population measures of location when using a robust estimator and there are missing values. For $J = 2$, methods have been studied based on trimmed means. But the methods are not readily extended to the case $J > 2$. Here, two alternative test statistics were considered, one of which performed poorly in some situations. The one method that performed well in simulations is based on a very simple test statistic with the null distribution approximated via a basic bootstrap technique. The method uses all of the available data to estimate each of the marginal (population) trimmed means. Other robust measures of location were considered, for which imputation methods have been derived, but in simulations the actual Type I error probability was estimated to be substantially less than the nominal level, even when there are no missing values.

Keywords Trimmed means, Minimum Covariance Determinant, OGK estimator, TBS estimator, Bootstrap methods, Imputation

1 Introduction

Consider $J \geq 2$ dependent groups and let θ_j ($j = 1, \dots, J$) be some robust measure of location associated with the j th marginal distribution. The paper considers the problem of testing

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_J \quad (1)$$

when there are missing values. A simple strategy is to use a complete case analysis. That is, exclude any rows of data where there are missing values and analyze the data that remain. But this approach might result in a reduction in power when testing hypotheses (e.g., [15]).

For the special case where θ_j is taken to be the population mean, various methods for handling missing observations have been derived and studied (e.g., [1, 20, 5, 29]). Approaches include a maximum likelihood method [12, 13], a weighted adjustment method [4],

single imputation [23] and multiple imputation [28, 16]. When testing (1), a simple strategy is to impute missing values and then use some conventional test statistic in the usual manner. It is known, however, that this approach can be unsatisfactory, as noted for example by Liang et al. [15] as well as Wang and Rao [32].

Ludbrook [18] suggested using a permutation test when comparing groups based on means. A positive feature of this permutation test is that when testing the hypothesis of identical distributions, the exact probability of a Type I error can be determined. But as a method for testing (1), it can be unsatisfactory even when there are no missing values and the goal is to compare measures of location (e.g., [2, 25]).

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample of size n from some bivariate distribution and let $D_i = X_i - Y_i$ ($i = 1, \dots, n$). Let μ_D be the corresponding population mean of D . An approach to missing values using an empirical likelihood method, based on single imputation, was derived by Liang et al. [15], which is readily adapted to the problem of computing a confidence interval for μ_D even when there are no missing values. But for heavy-tailed distributions, it can be unsatisfactory. For example, suppose D has the contaminated normal distribution

$$H(x) = (1 - \varepsilon)\Phi(x) + \varepsilon\Phi(x/10), \quad (2)$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. Based on the simulation with 5000 replications, if the sample size is $n = 100$, $\varepsilon = 0.1$, the actual Type I error is estimated to be 0.084 when testing at the .05 level. Using the Bartlett-correction studied by DiCiccio et al. [8], the actual probability coverage is 0.078. Here it was found that even with a few missing values, the probability coverage deteriorates, and so this approach was abandoned.

Even when there are no missing values, there are well known concerns about inferential methods based on means. A basic concern is that the population mean is not robust (e.g., [11, 30]). Moreover, the sample mean has a breakdown point of only $1/n$, roughly meaning that even a single outlier can result in the sample mean

being arbitrarily large or small. Also, outliers can result in very poor power relative to methods based on some robust measure of location (e.g., [34]). Indeed, even a small departure from normality can result in a relatively large standard error and poor power.

One robust estimator that has been studied extensively is a 20% trimmed mean. A reason for considering 20% trimming, rather than other amounts of trimming, is that its efficiency compares well to the mean under normality, but unlike the mean, its standard error remains relatively small as we move toward more heavy-tailed distributions [26]. Consequently, under normality, power is nearly the same as hypothesis testing methods based on means. For $J = 2$, methods for handling missing values, when using a 20% trimmed mean, have been studied that perform well in simulations [33]. But it is unclear how these methods might be generalized to $J > 2$. (Arguments for considering other robust estimators can be made, but for the situation at hand they were found to be unsatisfactory for reasons outlined in Section 2.0.1.)

Here, two test statistics were considered coupled with several robust measures of location. Among these methods, only one performed well in simulations in terms of controlling the probability of a Type I error in a reasonably accurate manner. Details of this method are given in section 2 along with a brief outline of the methods that were considered and abandoned.

2 Methods That Were Considered

The one measure of location that performed well in simulations is the 20% trimmed mean. Momentarily consider a single random sample: X_1, \dots, X_n . The γ -trimmed mean is computed as follows. Let $X_{(1)} \leq \dots \leq X_{(n)}$ be the values written in ascending order and let $g = [n\gamma]$, $0 \leq \gamma < 0.5$, where $[n\gamma]$ is the greatest integer less than or equal to $n\gamma$. Then the γ -trimmed mean is

$$\bar{X}_t = \frac{1}{n - 2g} \sum_{i=g+1}^{n-g} X_{(i)}.$$

A 20% trimmed mean corresponds to $\gamma = 0.2$.

Now consider a random sample from some unknown J -variate distribution (X_{i1}, \dots, X_{iJ}) , $i = 1, \dots, n$. Rather than discard any row of data having one or more missing values, all of the data are used to compute each marginal trimmed mean. So if the j th column of data has no missing values, all n values are used even if there are missing values in some other column. If the j th column contains say m missing values, then the trimmed mean is based on the $n - m$ values that are available.

To test (1), where now θ is taken to be the population trimmed mean, the test statistic is

$$Q = \sum (\bar{X}_{tj} - \bar{X}_t)^2,$$

where $\bar{X}_t = \sum \bar{X}_{tj}/J$ and \bar{X}_{tj} is the trimmed mean for the j th group. The strategy is to approximate the null distribution of Q via a bootstrap method. In more detail, let $C_{ij} = X_{ij} - \hat{X}_{tj}$. That is, shift the empirical

distributions so that the null hypothesis is true. Next, a bootstrap sample is obtained by resampling, with replacement, n rows of data from the matrix

$$\begin{pmatrix} C_{11} & \cdots & C_{1J} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nJ} \end{pmatrix}$$

yielding

$$\begin{pmatrix} C_{11}^* & \cdots & C_{1J}^* \\ \vdots & & \vdots \\ C_{n1}^* & \cdots & C_{nJ}^* \end{pmatrix}.$$

For the j th column of the bootstrap sample just generated, compute \hat{X}_{tj}^* , the 20% trimmed mean. Note that all of the non-missing values are being used. Then compute

$$Q^* = \sum (\bar{X}_{tj}^* - \bar{X}_t^*)^2,$$

where $\bar{X}_t^* = \sum \bar{X}_{tj}^*/J$. Repeat this process B times yielding Q_1^*, \dots, Q_B^* . Put these B values in ascending order yielding $Q_{(1)}^* \leq \dots \leq Q_{(B)}^*$. Let α be the nominal Type I error probability and reject the hypothesis of equal 20% trimmed mean if $Q > Q_{(c)}^*$, where $c = (1 - \alpha)B$ rounded to the nearest integer.

Here, $B = 599$ was used, which has been found to perform well, in terms of controlling the Type I error probability, when testing other hypotheses based on some robust estimator [34]. But in terms of power, a larger choice for B might have practical value. Racine and MacKinnon [22] discuss this issue at length. (Also see [14].) Davidson and MacKinnon [7] proposed a pretest procedure for choosing B .

2.1 Alternative Methods That Performed Poorly

Although the trimmed mean used here guards against the deleterious effects of outliers among the marginal distributions, a possible criticism is that it does not deal with outliers in a manner that takes into account the overall structure of the data. There are several (affine equivariant) estimators that deal with this issue (e.g., Wilcox, 2012, section 6.3). For some of these estimators, imputation methods have been derived (e.g., [31, 6]). So a simple strategy would be to impute missing values and then use the test statistic Q in conjunction with the bootstrap method just described. However, even with no missing values, control over the Type I error probability was found to be poor. The estimators considered were the minimum covariance determinant estimator (see, for example, [27]), the Donoho and Gasko [9] trimmed mean, the orthogonal Gnanadesikan-Kettenring (OGK) estimator derived by Maronna and Zamar [19], and the translated-biweight S-estimator proposed by Rocke [24]. A basic concern is that when using Q and testing at the .05 level, the actual level can be less than .01 even with $n = 60$. (When using the method in [6], via the R package GSE, bootstrap samples were encountered where the measure of location could not be compute.) Consequently, further details are omitted.

An alternative hypothesis testing method, based on trimmed methods, was considered, details of which can be found in Wilcox (2012, p.393). Missing data were handled as previously indicated. But situations were found where the actual level was estimated to be much higher than the nominal level, so for brevity, further details are omitted.

3 Simulation Results

Simulations were used to study the finite sample properties of the method in Section 2. It is known that generally, both skewness and kurtosis can impact the probability of a Type I error when comparing measures of location, so data were generated where the marginal distributions have one of four distributions: normal, skewed and light-tailed, symmetric and heavy-tailed, and skewed and heavy-tailed. More precisely, a g -and- h distribution [10] was used to generate the data. To elaborate, let Z be a random variable generated from a standard normal distribution. Then

$$W = \begin{cases} \frac{\exp(gZ)-1}{g} \exp\left(\frac{hZ^2}{2}\right), & \text{if } g > 0 \\ Z \exp\left(\frac{hZ^2}{2}\right), & \text{if } g = 0 \end{cases}$$

has a g -and- h distribution. The parameters g and h determine the skewness and kurtosis of the distribution. (Data were generated with the R function `rmul` in the R package `WRS`.)

The four distributions considered here are: standard normal ($g = 0, h = 0$), asymmetric light-tailed ($g = 0.2, h = 0$), symmetric heavy-tailed ($g = 0, h = 0.2$), and asymmetric heavy-tailed ($g = 0.2, h = 0.2$). Table 1 summarizes some properties of the g -and- h distribution, where κ_1 indicates skewness and κ_2 indicates kurtosis.

Table 1. Some properties of the g -and- h distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.00
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

Here the sample size used was $n = 30$. Let m_j be the number of missing values in the j th group. For $J = 2, 4$ and 6 groups, the number of missing values was taken to be $(m_1, m_2) = (5, 5)$, $(m_1, m_2, m_3, m_4) = (5, 5, 5, 0)$ and $(m_1, m_2, m_3, m_4, m_5, m_6) = (5, 5, 5, 5, 5, 0)$, respectively. For the first group, the first m_1 values are taken to be missing. More precisely, for the first five rows of data, the values for variable 1 are missing but the values of the other variables are available. For rows 6-10, the values for the second group are missing, but the values for the other groups are observed, etc. Table 2 reports the estimated Type I error probabilities based on 2000 replications. Based on results in Pratt (1968), the hypothesis that the actual level is .05 would be rejected if the estimated level is less than or equal to .04 or greater than or equal to .06. Power, if the actual level is .03, is equal to .995. If the actual level is .07, power is .959.

Table 2. Estimated Type I error probabilities when there are missing values, $\alpha = 0.05, n = 30$

g	h	$J = 2$	$J = 4$	$J = 6$
0.0	0.0	0.067	.046	.041
0.2	0.0	0.066	.043	.038
0.0	0.2	0.056	.032	.024
0.2	0.2	0.055	.030	.023

Although the seriousness of a Type I error depends on the situation, Bradley [3] suggests that as a general guide, at a minimum the actual Type I error probability should be between 0.025 and 0.075 when testing at the .05 level. All indications are that this criterion is met among the situations considered when using the bootstrap method in Section 2 based on the test statistic Q .

4 Conclusions

As is evident, simulation results cannot guarantee that the bootstrap method considered here, used in conjunction with a 20% trimmed mean and the test statistic Q , will always provide reasonably accurate control over the the Type I error probability. The main result is that it is the only method that performed reasonably well among the situations and location estimators that were considered. Indeed, the other (affine equivariant) robust estimators that were considered did not perform well even with no missing values.

References

- [1] Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.
- [2] Boik, R. J. (1987). The Fisher-Pitman permutation test: A non-robust alternative to the normal theory F test when variances are heterogeneous. *British Journal of Mathematical and Statistical Psychology*, 40, 26–42.
- [3] Bradley, J. V. (1978) Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144–152.
- [4] Cochran, W. G. (1977). *Sampling Techniques* (3rd Ed.). New York: Wiley.
- [5] Daniels, M. J. & Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis* Boca Raton, FL: Chapman & Hall/CRC.
- [6] Danilov , M., Yohai, V. J. & Zamar, R. H. (2012): Robust Estimation of Multivariate Location and scatter in the presence of missing data. *Journal of the American Statistical Association*, 107, 1178–1186.

- [7] Davidson, R. & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews*, 19, 55–68.
- [8] DiCiccio, T., Hall, P. & Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Annals of Statistics*, 19, 1053–1061.
- [9] Donoho, D. L. & Gasko, M. (1992). Breakdown properties of the location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics*, 20, 1803–1827.
- [10] Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. Hoaglin, F. Mosteller & J. Tukey (Eds.) *Exploring data tables, trends, and shapes*. (pp. 461–515). New York: Wiley.
- [11] Huber, P. J. & Ronchetti, E. (2009). *Robust Statistics*, 2nd Ed. New York: Wiley.
- [12] Ibrahim, J. G., Chen, M. H. & Lipsitz, S. R. (1999). Monte Carlo EM for missing covariates in parametric regression models. *Biometrics*, 55, 591–596.
- [13] Ibrahim, J. G., Lipsitz, S. R. & Horton, N. (2001). Using auxiliary data for parameter estimation with nonignorable missing outcomes. *Applied Statistics*, 50, 361–373.
- [14] Jöckel, K.-H. (1986). Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Annals of Statistics*, 14, 336–347.
- [15] Liang, H., Su, H. & Zou, G. (2008). Confidence intervals for a common mean with missing data with applications in an AIDS study. *Computational Statistics & Data Analysis*, 53, 546–553.
- [16] Little, R. J. A. & Rubin, D. (2002). *Statistical Analysis with Missing Data*, 2nd Ed. New York: Wiley.
- [17] Liu, R. G. & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, 92, 266–277.
- [18] Ludbrook, J. (2008). Outlying observations and missing values: how should they be handled? *Clinical and Experimental Pharmacology & Physiology*, 35, 670–678.
- [19] Maronna, R. A. & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional data sets. *Technometrics*, 44, 307–317
- [20] McKnight, P. E., McKnight, K. M., Sidani, S. & Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. New York: Guilford Press.
- [21] Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. *Journal of the American Statistical Association*, 63, 1457–1483.
- [22] Racine, J. & MacKinnon, J. G. (2007). Simulation-based tests than can use any number of simulations. *Communications in Statistics-Simulation and Computation*, 36, 357–365.
- [23] Rao, J.N.K. and Sitter, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, 453-460.
- [24] Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics*, 24, 1327–1345.
- [25] Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association* 85, 686–692.
- [26] Rosenberger, J. L. & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. Hoaglin, F. Mosteller and J. Tukey (Eds.) *Understanding Robust and exploratory data analysis*. (pp. 297–336). New York: Wiley.
- [27] Rousseeuw, P. J. & van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–223.
- [28] Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- [29] Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall
- [30] Staudte, R. G. & Sheather, S. J. (1990). *Robust Estimation and Testing* New York: Wiley.
- [31] Todorov, V., Templ, M. & Filzmoser, P. (2011). Detection of multivariate outliers in business survey data with incomplete information. *Advances in Data Analysis and Classification*, 5, 37–56.
- [32] Wang, Q.H. & Rao, J. N. K. (2002). Empirical likelihood-based inference in linear models with missing data. *Scandinavian Journal of Statistics*, 29, 563–576.
- [33] Wilcox, R. R. (2011). Comparing two dependent groups: Dealing with missing values *Journal of Data Science*, 9, 1–13.
- [34] Wilcox, R. R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*, 3rd Ed. San Diego, CA: Academic Press.