Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 6

11-1-2016

Within Groups ANOVA When Using a Robust Multivariate Measure of Location

Rand Wilcox
University of Southern California, rwilcox@usc.edu

Timothy Hayes
University of Southern California

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, 41-52.
doi: 10.22237/jmasm/1478001840

ISSN 1538 − 9472

Within Groups ANOVA When Using a Robust Multivariate Measure of Location

Rand Wilcox
University of Southern California
Los Angeles, California

Timothy Hayes
University of Southern California
Los Angeles, California

For robust measures of location associated with *J* dependent groups, various methods have been proposed that are aimed at testing the global hypothesis of a common measure of location applied to the marginal distributions. A criticism of these methods is that they do not deal with outliers in a manner that takes into account the overall structure of the data. Location estimators have been derived that deal with outliers in this manner, but evidently there are no simulation results regarding how well they perform when the goal is to test the some global hypothesis. The paper compares four bootstrap methods in terms of their ability to control the Type I error probability when the sample size is small, two of which were found to perform poorly. The choice of location estimator was found to be important as well. Indeed, for several of the estimators considered here, control over the Type I error probability was very poor. Only one estimator performed well when using the first of two general approaches that might be used. It is based on a variation of the (affine equivariant) Donoho-Gasko trimmed mean. For the second general approach, only a skipped estimator performed reasonably well. (It removes outliers via a projection method and averages the remaining data.) Only one bootstrap method was found to perform well when using the first approach. A different bootstrap method is recommended when using the second approach.

Keywords: Bootstrap methods, outliers, skipped estimator, Donoho-Gasko trimmed mean

Introduction

Methods for comparing dependent groups, based on the usual sample mean, are not robust under general conditions. A fundamental concern with any inferential technique based on the mean is that it can result in relatively low power when dealing with heavy-tailed distributions (e.g., Marrona, Martin, & Yohai, 2006; Staudte & Sheather, 1990; Wilcox, 2012). Roughly, heavy-tailed distributions are

Rand Wilcox is an Professor in the Department of Psychology. Email him at: rwilcox@usc.edu. Timothy Hayes is a doctoral candidate in the Psychology Department.

characterized by outliers that inflate the standard error of the sample mean. Even an arbitrarily small departure from normality can result in poor power. Another concern is that the breakdown point of the sample mean is only 1 / *n*, where n is the sample size. That is, the minimum proportion of points that must be altered to completely destroy the sample mean (make it arbitrarily large or small) is 1 / *n*.

Various methods for comparing $J \geq 2$ dependent groups have been derived and studied that are based on replacing the marginal means with some robust estimator (e.g., Wilcox, 2012, Ch. 8). That is, if X_{ij} ($i = 1, \ldots, n$; $j = 1, \ldots, J$) is a random sample of *n* vectors from some *J*-variate distribution, for each *j*, a robust measure of location is computed. These methods deal with outliers among the marginal distributions, but they do not deal with outliers in a manner that takes into account the overall structure of the data. As a simple example of what this means, it is not unusual to be young, it is not unusual to have heart disease, but it is very unusual to be both young and have heart disease.

Situations are encountered where there are no outliers among the marginal distributions based on, for example, a boxplot or the MAD-median rule, yet there are outliers when using a multivariate outlier detection technique that takes into account the overall structure (e.g., Wilcox, 2012).

Another possible criticism of applying a robust estimator to each of the marginal distributions is that the resulting measure of location is not affine equivariant (e.g., Rousseeuw & Leroy, 1987). To elaborate, note that a basic requirement for $\hat{\theta}_j$ to qualify as a location estimator is that it be both scale and location equivariant. That is, if $\hat{\theta}_j = T(X_{ij}, \ldots, X_{nj})$ is some estimate of θ_j, then for $\hat{\theta}_j$ to qualify as a location estimator, it should be the case that for constants *a* and *b*,

$$T\left(aX_1 + b, \ldots, X_n A + b\right) = aT\left(X_1, \ldots, X_n\right) + b.$$

In the multivariate case, a generalization of this requirement, affine equivariance, is that for a *J*-by-*J* nonsingular matrix **A** and vector **b** having length *J*,

$$T\left(\mathbf{X}_1\mathbf{A} + \mathbf{b}, \ldots, \mathbf{X}_n\mathbf{A} + \mathbf{b}\right) = T\left(\mathbf{X}_1, \ldots, \mathbf{X}_n\right)\mathbf{A} + \mathbf{b}. \quad (1)$$

In particular, the estimate is transformed properly under rotations of the data as well as changes in location and scale.

The goal in this paper is to report simulation results on several methods for comparing dependent groups with an emphasis on situations where the sample size is small. Several multivariate estimators were considered that take into account the overall structure of the data when dealing with outliers. All of them are location and scale equivariant, but one is not affine equivariant.

Here, two types of global hypotheses are considered. To describe them, let $\hat{\Theta}(\mathbf{X})$ represent one of the multivariate location estimators to be considered. Letting $\Theta = (\theta_1, \ldots, \theta_j)$ represent the estimand associated with $\hat{\Theta}(\mathbf{X})$ (the population analog of $\hat{\Theta}(\mathbf{X})$, the first global hypothesis is

$$H_0 : \theta_1 = \cdots = \theta_j \tag{2}$$

To describe the second hypothesis, let $D_{ijk} = X_{ij} - X_{ik}$, $j < k$, and let $\hat{\Theta}(D)$ be some multivariate location estimator based on the D_{ijk} values. There are $L = (J^2 - J)/2$ parameters being estimated, which are labeled $\Delta = (\delta_1, ..., \delta_L)$. Now the goal is to test

$$H_0 : \delta_1 = \cdots = \delta_L = 0. \tag{3}$$

From basic principles, when dealing with means, there is no distinction between (2) and (3). But under general conditions, this is not the case when using a robust estimator. (It is readily verified, for example, that the difference between the marginal medians is not necessarily equal to the median of the difference scores.)

Two bootstrap methods for testing (2) were considered here, and another two methods were considered when testing (3). As will be seen, the choice of estimator, as well as the bootstrap method that is used, is crucial in terms of controlling the Type I error probability, at least when the sample size is small.

Description of the Methods

The Location Estimators

The first estimator is based on a particular variation of an affine equivariant estimator derived by Donoho and Gasko (1992), which will be labeled the DG estimator henceforth. Roughly, it begins by quantifying how deeply each point is

nested within the cloud of points. Here, this is done using a projection-type method, which provides an approximation of half-space depth (Wilcox, 2012, section 6.2.5). To elaborate, let $\hat{\tau}$ be some initial affine equivariant location estimator. Here, the (fast) minimum covariance determinant estimator (MCD) is used (e.g., Wilcox, 2012, section 6.3.2). Briefly, the MCD estimator searches for a subset of half the data that minimizes the generalized variance. The mean of this subset is the MCD measure of location. Let

$$\mathbf{U}_i = \mathbf{X}_i - \hat{\tau}$$

$$B_i = \mathbf{U}_i \mathbf{U}_i'$$

(i = 1, …, n) and for any j (j = 1, …, n), let

$$W_{ij} = \sum_{k=1}^{j} U_{ij} U_{jk}$$

and

$$T_{ij} = \frac{W_{ij}}{B_i}\left(U_{i1}, \ldots, U_{ij}\right) \tag{4}$$

The distance between $\hat{\theta}$ and the projection of X_j (when projecting onto the line connecting $\mathbf{X}_i$ and $\hat{\tau}$) is

$$H_{ij} = \text{sign}\left(W_{ij}\right)\left\|T_{ij}\right\|$$

where $\left\|T_{ij}\right\|$ is the Euclidean norm associated with the vector T_{ij}.

Let d_{ij} be the depth of X_j when projecting points onto the line connecting X_i and $\hat{\theta}$. That is, for fixed i and j, the depth of the projected value of X_j is

$$d_{ij} = \min\left(\#\left\{H_{ij} \le H_{ik}\right\}, \#\left\{H_{ij} \ge H_{ik}\right\}\right),$$

Where $\#\{H_{ij} \le H_{ik}\}$ indicates how many H_{ik} (k = 1, …, n) values satisfy $H_{ij} \le H_{ik}$. The depth of X_j is taken to be $L_j = \min d_{ij}$, the minimum being taken over all i = 1, …, n.

The Donoho-Gasko (DG) γ trimmed mean associated with the X_{ij} values is the average of all points that are at least γ deep in the sample. That is, points having depth less than γ are trimmed and the mean of the remaining points is computed. If the maximum depth among all n points is at least γ, the breakdown point of the DG estimator is $\gamma / (1 + \gamma)$, where the breakdown point refers to the minimum proportion of points that must be altered to completely destroy an estimator. Here, $\gamma = .2$ is used.

The other estimator considered here, which performed well in simulations when testing (3), is a skipped estimator based on a projection method for detecting outliers, which will be labeled the SP estimator. Fix i, and for the point $\mathbf{X}_i$ let

$$\mathbf{A}_i' = \mathbf{X}_i - \hat{\tau},$$

$$\mathbf{B}_j' = \mathbf{X}_i - \hat{\tau}$$

$$\mathbf{C}_j = \frac{A_i B_j}{B_j B_j'} \mathbf{B}_j,$$

$j = 1, \ldots, n$. Then when projecting the points onto the line between $\mathbf{X}_i$ and $\hat{\tau}$, the distance of the j^{th} point from $\hat{\tau}$ is

$$V_{ij} = \|C_j\|.$$

The j^{th} point is declared an outlier if

$$V_{ij} > M_V + c(q_2 - q_1) \tag{5}$$

Where M_V, q_1 and q_2 are the usual sample median and estimates of the lower and upper quartiles, respectively, based on the $V_{i1}, \ldots, V_{in}$ values, and c is the .95 quantile of a chi-squared distribution with J degrees of freedom. (Here, the quartiles are estimated via the ideal fourths; see Frigge, Hoaglin, & Iglewicz, 1989.)

The process just described is for a single projection. Repeating this process for each i ($i = 1, \ldots, n$), $\mathbf{X}_j$ is declared an outlier if for any of these projections, V_{ij} satisfies (5). Removing any points declared an outlier, the mean of the remaining

data is taken to be the SP estimator of location. Its small-sample efficiency compares well to the DG estimator (Wilcox, 2012). Note that the estimate of interquartile range, $q_2 - q_1$, based on the ideal fourths, has a breakdown point of .25 indicating that the breakdown point of the SP estimator is .25 as well. The small-sample efficiency of the SP estimator compares well to several other robust estimators that have been derived (Ng & Wilcox, 2010).

Several other affine equivariant estimators were considered but which performed poorly in simulations in terms of controlling the Type I error probability. So computational details related to these other estimators are not provided. They included the minimum volume ellipsoid (MVE) estimator (Rousseeuw & van Zomeren, 1990), the minimum covariance determinant (MCD) estimator (Rousseeuw & Van Driessen, 1999), the translated-biweight S-estimator (Rocke, 1996), the median ball algorithm (Olive, 2004) and the orthogonal Gnanadesikan-Kettenring (OGK) estimator (Maronna & Zamar, 2002).

Testing (2) and (3)

Two bootstrap methods for testing (2), as well as (3), were considered. The first, which is designed to test (2) and corresponds to the RMPB3 in Wilcox (2012, section 8.2.5), is applied as follows. Compute the test statistic

$$Q = \sum \left(\hat{\theta}_j - \bar{\theta} \right)^2,$$

Where $\bar{\theta} = \sum \hat{\theta} / J$. An appropriate critical value is estimated by first setting $Z_{ij} = X_{ij} - \hat{\theta}_j$. 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 from the matrix $\mathbf{Z}$ yielding $Z_{ij}^* \left(i = 1, \ldots, n;\ j = 1, \ldots, J \right)$. Compute the measure of location that is of interest based on this bootstrap sample yielding $\hat{\theta}_j^*$ and test statistic Q^*. Repeat this process B times yielding $Q_1^*, \ldots, Q_B^*$. Put these B values in ascending order yielding $Q_{(1)}^* \leq \cdots \leq Q_{(B)}^*$. Then reject the hypothesis of equal measures of location at the α level if $Q > Q_{(u)}^*$, where $u = (1 - \alpha) B$ rounded to the nearest integer.

The second method for testing (2) is based in part on bootstrap samples obtained from the X_{ij} values rather than the Z_{ij} values. The strategy is based on determining how deeply the grand mean is nested within the resulting bootstrap

cloud. Details about this strategy can be found in Wilcox (2012, pp. 392-393). Because this approach performed poorly for the situation at hand, no details are provided.

The two bootstrap methods for testing (3) can be roughly described as follows. Take B bootstrap samples by resampling with replacement from the matrix **X**, compute a measure of location based on the resulting difference scores and determine how deeply the null vector 0 is nested within the bootstrap cloud. Here, two methods were used to measure the depth of a point in data cloud: Mahalanobis distance and projection distance. In general this approach did not perform well. But when coupled with the DG estimator, it did perform reasonably well when testing (3).

To provide more details, let $\hat{\Delta}_b^*\left(b=1,\ldots,B\right)$ indicate the location estimate of Δ based on the b^{th} bootstrap sample and for convenience let $\hat{\Delta}_0^*$ denote the null vector. Let $P_d^*\left(\hat{\Delta}_b^*\right)$ be the projection distance of $\hat{\Delta}_b^*$ based on the $B+1$ points $\hat{\Delta}_0^*,\ldots,\hat{\Delta}_B^*$. So $P_d^*\left(\hat{\Delta}_b^*\right)$ reflects how far the null vector is from the center of the bootstrap cloud. Then, from general theoretical results in Liu and Singh (1997), a p-value is

$$1-\frac{1}{B}\sum_{b=1}^{B} I\left(P_d^*\left(\hat{\Delta}_0^*\right)\geq P_d^*\left(\hat{\Delta}_b^*\right)\right)$$

where the indicator function $I\left(P_d^*\left(\hat{\Delta}_0^*\right)\geq P_d^*\left(\hat{\Delta}_b^*\right)\right)=1$ if $P_d^*\left(\hat{\Delta}_0^*\right)\geq P_d^*\left(\hat{\Delta}_b^*\right)$; otherwise $I\left(P_d^*\left(\hat{\Delta}_0^*\right)\geq P_d^*\left(\hat{\Delta}_b^*\right)\right)=0$. This will be called method D-P. When the projection distance is replaced by Mahalanobis distance, this will be called method D-M.

Simulation

Simulations were used to study the small-sample properties of the methods described in the previous section. The simulations were run using the software R, with much of the code written in C++. In addition, the R functions took advantage of a multi-core processor via the R package parallel. Despite this, execution time was relatively high, particularly when using the DG estimator in conjunction with method D-P. Consequently, estimated Type I error probabilities were based on 2000 replications. Four types of distributions were used: normal, symmetric and

heavy-tailed, asymmetric and light-tailed, and asymmetric and heavy-tailed. More precisely, the marginal distributions were taken to be one of four *g*-and-*h* distributions (Hoaglin, 1985) that contain the standard normal distribution as a special case. (The R function rmul, in Wilcox, 2012, was used to generate observations.) If Z has a standard normal distribution and $g > 0$, then

$$W = \frac{\exp(gZ) - 1}{g} \exp(hZ^2/2)$$

has a *g*-and-*h* distribution where g and h are parameters that determine the first four moments. If $g = 0$, this last equation is taken to be

$$W = Z \exp(hZ^2/2).$$

The four distributions used here were the standard normal ($g = h = 0.0$), asymmetric heavy-tailed distribution ($h = 0.2$, $g = 0.0$), an asymmetric distribution with relatively light tails ($h = 0.0$, $g = 0.2$), and an asymmetric distribution with heavy tails ($g = h = 0.2$). Table 1 shows the skewness (κ_1) and kurtosis (κ_2) for each distribution. Additional properties of the *g*-and-h distribution are summarized by Hoaglin (1985). The number of bootstrap samples was taken to be $B = 500$. This choice generally seems to perform well in other settings, in terms of controlling the Type I error probability (Wilcox, 2012). But a possibility is that a larger choice for B might yield more power (e.g., Racine & MacKinnon, 2000). The correlation among the variables was taken to be $\rho = 0$ or $\rho = .5$.

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

g	*h*	κ_1	κ_2
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

As a partial check on the impact of heteroscedasticity on the Type I error probability, the X_{ij} values were taken to be λX_{ij} ($i = 1, \ldots, n$). The two choices for λ were 1 and 4. For symmetric *g*-and-*h* distributions ($g = 0$), all of the measures of location considered here are equal to zero, so for $\lambda = 4$ the null hypothesis remains

true. But when dealing with skewed distributions ($g > 0$), this is not the case. To deal with this, the expected value of an estimator was determined by generating 4000 samples of size n from a specified g-and-h distribution (with $\lambda = 1$) and then averaging the resulting estimates. So with $p = 4$, in essence 16,000 estimates are being used. Then the marginal distributions were shifted so that, based on the expected value of an estimator, the null hypothesis is true.

Shown in Table 2 are the results when using the SP estimator with methods D-M and D-P to test (3). Although the seriousness of a Type I error depends on the situation, Bradley (1978) has suggested that as a general guide, when testing at the .05 level, at a minimum the actual level should be between .025 and .075. As can be seen, this criterion is generally met when using D-M. But under normality, with $\rho = .5$, is this not the case, the largest estimate being .098. In contrast, when using D-P, the largest estimate is .075.

Table 2. Estimated Type I error probabilities when testing (3), $n = 20$, $\alpha = .05$ using the SP estimator.

		D-M				D-P			
		$\lambda = 1$		$\lambda = 4$		$\lambda = 1$		$\lambda = 4$	
g	h	$\rho = .0$	$\rho = .5$	$\rho = .0$	$\rho = .5$	$\rho = .0$	$\rho = .5$	$\rho = .0$	$\rho = .5$
0.0	0.0	.069	.065	.096	.083	.055	.063	.075	.065
0.0	0.2	.052	.047	.055	.049	.033	.042	.041	.043
0.2	0.0	.070	.071	.039	.046	.054	.070	.054	.056
0.2	0.2	.044	.044	.030	.040	.035	.039	.028	.040

Reported in Table 3 are simulation results when using method Q to test (2) with the DG estimator and $n = 30$. For $n = 20$, estimated Type I error probabilities exceed .075. But as indicated in Table 3, with $n = 30$, the estimates ranged between .025 and .061 when testing at the .05 level. When testing (2) instead via methods D-M or D-P, control over the Type I error probability was poor.

Table 3. Estimated Type I error probabilities, $n = 30$, $\alpha = .05$ using method Q to test (2) with the DG estimator

		$\lambda = 1$		$\lambda = 4$	
g	h	$\rho = .0$	$\rho = .5$	$\rho = .0$	$\rho = .5$
0.0	0.0	.056	.057	.053	.060
0.0	0.2	.031	.034	.040	.041
0.2	0.0	.054	.060	.057	.061
0.2	0.2	.026	.025	.038	.040

Conclusion

When using a location estimator that takes into account the overall structure of data when dealing outliers, finding a method for testing (2) and (3) appears to be nontrivial when the sample size is small. The bulk of the methods considered here performed poorly in terms of controlling the Type I error probability, particularly when using an affine equivariant estimator.

Only one method performed well in simulations when testing (2) and an affine equivariant estimator is used: method *Q* in conjunction with the DG estimator. No method based on an affine equivariant estimator was found to perform reasonably well when testing (3). Moreover, several bootstrap methods that perform reasonably well using a robust estimator applied to each of marginal distributions did not perform well for the situations considered here. However, the skipped estimator studied here, which is location and scale equivariant, was found to perform reasonably well when testing (3) via a percentile bootstrap method that measures the depth of null vector using projection distances. Another possible appeal of the SP estimator over the DG estimator is that for light-tailed distributions, including normal distributions, the DG estimator has relatively poor efficiency (e.g., Massé & Plante, 2003; Wilcox, 2012, p. 251). In contrast, the SP estimator performs nearly as well as the usual sample mean.

R functions are available for applying the methods that performed well in the simulations. The R function bd1GLOB tests (2). The DG estimator can be used by setting the argument est=dmean. Setting the argument MC=TRUE takes advantage of multi-core processor, if multiple cores are available, via the R package parallel, which can be installed via R command install.packages. The R function rmdzD applies method D-P in conjunction with the SP estimator. Again, setting the argument MC=TRUE will take advantage of a multi-core processor if one is available and the R package parallel has been installed. These functions can be installed with the R command install.packages(``WRS",repos=``http:R-Forge.R-project.org"). They are also stored in the file Rallfun-v24, which can be downloaded from the first author's web page.

References

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x

Donoho, D. L. & Gasko, M. (1992). Breakdown properties of the location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics*, *20*(4), 1803-1827. doi: 10.1214/aos/1176348890

Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some implementations of the boxplot. *American Statistician*, *43*(1), 50-54. doi: 10.1080/00031305.1989.10475612

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.) *Exploring data tables trends and shapes* (pp. 461-515). Hoboken, NJ: John Wiley & Sons, Inc.

Liu, R. Y. & Singh, K. (1997). Notions of limiting *p* values based on data depth and bootstrap. *Journal of the American Statistical Association*, *92*(437), 266-277. doi: 10.1080/01621459.1997.10473624

Maronna, R. A., Martin, D. R., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. New York, NY: Wiley.

Maronna, R. A., & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional data sets. *Technometrics*, *44*(4), 307-317. doi: 10.1198/004017002188618509

Massé, J.-C., & Plante, J.-F. (2003). A Monte Carlo study of the accuracy and robustness of ten bivariate location estimators. *Computational Statistics & Data Analysis*, *42*(1/2), 1-26. doi: 10.1016/S0167-9473(02)00103-2

Ng, M., & Wilcox, R. R. (2010). The small-sample efficiency of some recently proposed multivariate measures of location. *Journal of Modern and Applied Statistical Methods*, *9*(1), 28-42. Available at: http://digitalcommons.wayne.edu/jmasm/vol9/iss1/5

Olive, D. J. (2004). A resistant estimator of multivariate location and dispersion. *Computational Statistics & Data Analysis*, *46*(1), 93-102. doi: 10.1016/S0167-9473(03)00119-1

Racine, J. S., & MacKinnon, J. G. (2007). Simulation-based tests that can use any number of simulations. *Communications in Statistics - Simulation and Computation*, *36*(2), 357-365. doi: 10.1080/03610910601161256

Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics*, 24(3), 1327-1345. doi: 10.1214/aos/1032526972

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York, NY: John Wiley & Sons, Inc.

Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics, 41*(3), 212-223. doi: 10.1080/00401706.1999.10485670

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association, 85*(411), 633-639. doi: 10.2307/2289995

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. New York, NY: Wiley.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). San Diego, CA: Academic Press.