

# Inadvertent Use of ANOVA in Educational Research: ANOVA is not A Surrogate for MANOVA \*

Lokman AKBAY \*\* Tuncer AKBAY \*\*\* Osman EROL \*\*\*\* Mustafa KILINÇ \*\*\*\*\*

## Abstract

ANOVA and MANOVA address different research questions and decision on conducting one or the other of these tests relies on the research purpose. One prominent illegitimate analysis of multivariate data is developed out of conducting multiple ANOVAs rather than conducting a MANOVA. Another common mistake about MANOVA applications is the use of improper post hoc procedure. Post hoc procedures are needed to determine why the null hypothesis was rejected. Although the correct post hoc procedure for MANOVA is descriptive discriminant analysis (DDA), many researchers fail to conduct DDA to interpret their MANOVA results. The purpose of this study is two-fold; (1) we aim to emphasize the theory behind the MANOVA and its appropriate post hoc procedure and make clear distinction between surrogate statistical procedures such as ANOVA; and (2) this study also investigates the extent of incorrect analysis of multivariate dependent variables in educational research in Turkey. First, we provided a small simulation study to demonstrate the extent to which multiple ANOVAs yields contradictory results when they are inadvertently used to test group mean differences on multiple dependent variables. Results of the simulations indicated that MANOVA and multiple ANOVAs had severe disagreements under many conditions. Disagreement rate is elevated under the conditions where MANOVA retains the null hypothesis. Then, we systematically reviewed the archives of three education journals, which are classified as higher-, medium, and lower quality journals. Results indicated that correct use of MANOVA with its proper post hoc procedure is not common practice across educational researchers who publish in Turkish education journals.

*Key Words:* Multivariate data analysis, multivariate dependent variable, ANOVA, MANOVA.

## INTRODUCTION

Univariate and multivariate data analysis are the two distinct statistical approaches. Univariate analysis involves only one variable at a time while two or more variables are involved in multivariate analysis. The analysis on group mean differences on a single outcome variable is referred to as Analysis of Variance (ANOVA); yet when multiple outcome variables are involved, we speak of Multivariate Analysis of Variance (MANOVA) (Fish, 1988; Stevens, 2002). Primary purpose of conducting both analyses is to determine treatment variable effect. MANOVA can be considered as a more general procedure of ANOVA. Although MANOVA is the most commonly used multivariate data analysis procedure (Kieffer, Reese & Thompson, 2001; Zientek & Thompson, 2009); literature indicates that MANOVA and its accompanying post hoc procedures are not properly understood by a considerable amount of social science researchers (Tonidandel & LeBreton, 2013; Warne, 2014; Warne, Lazo, Ramos & Ritter, 2012).

\* Preliminary results of this work were presented at the 27<sup>th</sup> International Conference on Educational Sciences, Antalya, Turkey 2016.

\*\* Assist. Prof., Burdur Mehmet Akif Ersoy University, Faculty of Education, Burdur-Turkey, lokmanakbay@gmail.com, ORCID ID: 0000-0003-4026-5241

\*\*\* Ph.D., Burdur Mehmet Akif Ersoy University, Faculty of Education, Burdur-Turkey, tuncerakbay@mehmetakif.edu.tr, ORCID ID: 0000-0003-3938-1026

\*\*\*\* Assist. Prof. Burdur Mehmet Akif Ersoy University, Faculty of Education, Burdur-Turkey, oerol@mehmetakif.edu.tr, ORCID ID: 0000-0002-9920-5211

\*\*\*\*\* Assist. Prof., Burdur Mehmet Akif Ersoy University, Faculty of Education, Burdur-Turkey, mkilinc@mehmetakif.edu.tr, ORCID ID: 0000-0002-2759-4916

To cite this article:

Akbay, L., Akbay, T., Erol, O., & Kılınç, M. (2019). Inadvertent use of ANOVA in educational research: ANOVA is not a surrogate for MANOVA. *Journal of Measurement and Evaluation in Education and Psychology*, 10(3), 302-314. doi: 10.21031/epod.524511

Received: 08.02.2019

Accepted: 30.05.2019

ANOVA and MANOVA address different research questions so that decision on conducting one or the other of these analyses must be determined by the purpose of the research. One prominent inadvertent analysis of multivariate data is derived from conducting multiple ANOVAs rather than conducting a MANOVA. Conducting multiple ANOVAs fundamentally differs from MANOVA in two ways: (1) Multiple ANOVAs yield increase in the likelihood of committing Type I error. In a series of ANOVA, experiment-wise error can be as high as  $1-(1-\alpha)^t$ , where  $\alpha$  is the Type I error rate and  $t$  is the number of ANOVAs conducted. For instance, the experiment-wise error will be .185 (i.e.,  $1-(1-.05)^4$ ) for  $\alpha = .05$  and  $t = 4$ . Of course, this is the extreme case where dependent variables are uncorrelated. It should be noted that Type I error rate inflation depends on the correlation between the dependent variables (Hummel & Sligo, 1971). Therefore, Bonferroni correction (i.e.,  $\alpha/t$ ) cannot overcome this problem unless dependent variables are truly uncorrelated.

Second fundamental difference (2) relies on the fact that ANOVA and MANOVA tend to answer to distinct empirical questions. Former statistical procedure is used to test the group mean differences on an observed variable, whereas the latter is used to test the group mean differences on underlying latent variables (Zientek & Thompson, 2009). Multiple ANOVAs fail to determine relationship between the independent variable(s) and combination of dependent variables (Warne, 2014). Notice that we are not interested in the possible group mean difference on indicators (i.e., observed variables) of a latent dependent variable; yet we would like to detect the group mean difference on the latent variable that may be determined by a linear combination of the indicator variables. For example, from the statistical point of view, there might be no statistically significant difference in each of the dependent variables, yet a significant difference might be suggested by combination of them.

Another common mistake that is made in conducting MANOVA is related to use of improper post hoc procedure. Post hoc procedures are generally needed when the null hypothesis ( $H_0$ ) is rejected in MANOVA (Stevens, 2002) to determine why the  $H_0$  was rejected. Although the proper post hoc procedure for MANOVA is descriptive discriminant analysis (DDA) (Warne, 2014), most researchers do not conduct DDA to interpret their MANOVA results (Huberty & Morris, 1989; Warne et al., 2012). This is mainly because many researchers use SPSS for MANOVA and it automatically conducts an ANOVA for each dependent variable. However, some researchers claim that because ANOVA is only concerned with observed variable, use of ANOVA as a follow-up procedure to significant MANOVA result is against the nature of MANOVA (Kieffer et al., 2001; Zientek & Thompson, 2009). Underlying rationale to this claim relies on the difference in the empirical questions ANOVA and MANOVA are exposed to (i.e., ANOVA tests the mean differences on the observed variable whereas MANOVA tests the mean differences on the underlying latent variables).

### ***Purpose of the Study***

The purpose of this study is two-fold. (1) We aim to emphasize the theory behind the MANOVA and its appropriate post hoc procedure (i.e., DDA) and make clear distinction between surrogate statistical procedures such as ANOVA. (2) This study also investigates the extent of inadvertent analysis of multivariate dependent variables in educational research in Turkey. In other words, this study aims to determine to the extent to which educational researchers conduct MANOVA when it is the most appropriate way of analyzing the data to answer their empirical question.

### ***Univariate and Multivariate Hypothesis Testing***

To find out whether the mean score on a dependent variable is equal across two or more groups, ANOVA test is conducted and an F-statistic is computed. To test the null hypothesis (i.e., group means are equal) observed F-statistic compared against the sampling distribution. The null hypothesis is rejected when observed statistic fall beyond a predetermined critical value; otherwise the null hypothesis is retained. When multiple dependent variables are employed in the analysis, each of them may or may not fall in the rejection region. Furthermore, linear combinations of the dependent variables may or may not fall in the rejection region. Imagine a case where two perfectly uncorrelated

dependent variables are tested; as can be seen in Figure 1, rejection region becomes the outside of the circle. Further assume that these two uncorrelated observed dependent variables equally contribute to the underlying latent variable. Then one of the four possible cases may be observed.

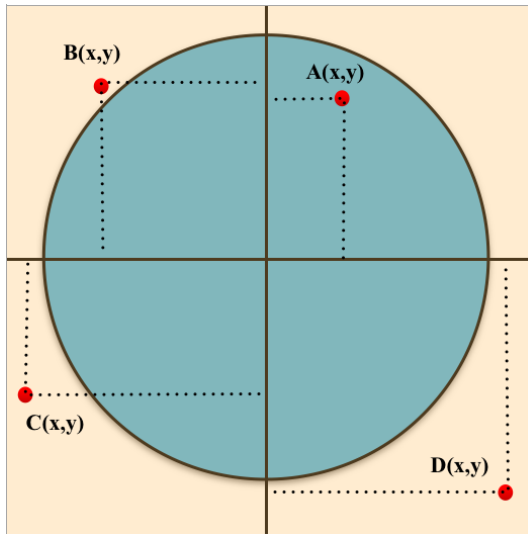


Figure 1. Possible Hypothesis Testing Results for Two Perfectly Uncorrelated Dependent Variables

In the first case, both of the observed variables (i.e.,  $x$  and  $y$ ) and the latent variable (i.e.,  $A$ ) do not fall outside the circle so that all of the hypotheses testing the group mean differences in the observed and latent variables are retained. In other words, neither the ANOVAs nor the MANOVA suggest any significant difference. In the second case (i.e., latent variable  $B$ ), although both ANOVAs fail to reject the null hypotheses, MANOVA rejects the null hypothesis. In case of latent variable  $C$ , MANOVA and ANOVA testing the difference on observed variable  $x$  yield significant difference; whereas ANOVA for the observed variable  $y$  suggests no significant difference. In the last case, all tests reject the null hypotheses. As shown in the Figure 1, MANOVA and multiple ANOVAs may result in contradicting results.

When a MANOVA test result rejects the null hypothesis of equality of group means we need to go ahead and identify how one or more groups of observations differ by interrelated multiple dependent variables. Difference can be in anywhere: in one variable or in a combination of multiple variables. DDA should be run to find the source of the difference. Although we have no intention to explain DDA in details, several reminders might be noted here. DDA provides us with discriminant functions, which are created by the linear combination of the dependent variables to maximize group differences (Sherry, 2006). DDA treats outcome variables as the linear combination of the dependent variables that maximizes group differences. DDA, in general, help us determine how much each of the dependent variable contribute to group difference on the outcome variable.

## METHOD

This study can be regarded a documentary survey, which is a type of survey research under the descriptive research method. Documentary surveys are akin to content analysis or document analysis. The term content analysis is used to define the process of summarizing and reporting written data (Cohen, Manion & Morrison, 2002). Document analysis is defined as a systematic procedure for evaluating or reviewing printed and/or electronic materials (Bowen, 2009). With this documentary survey, we aim to ascertain whether use of MANOVA with its proper post hoc procedure is common practice across educational researchers who publish in Turkish education journals.

### Data Collection Procedure

We have obtained our data through screening the archives of three education journals. We specifically reviewed all issues published in these journals in the last four years (i.e., 2015-2018). These three journals may represent the higher, medium, and lower quality journals based on where they are indexed. Based on our classification, Thompson Reuters Social Sciences Citation Index indexes the higher quality journal. The medium and lower quality journals are indexed by the Thompson Reuters Emerging Sources Citation index and ULAKBIM Social and Human Sciences Database (Sosyal ve Beşeri Bilimler Veri Tabanı), respectively. Detailed information on the journals may be provided upon request.

Researchers reviewed the articles published in these three Turkish education journals and reported the counts as well as the type of analyses used to test intervention effect or group mean differences in multiple outcome variables. Furthermore, counts and the types of post-hoc procedures are also reported. We considered the following types of multivariate dependent variable analyses:

1. Use of MANOVA to test group mean differences in multivariate data
  - followed by DDA
  - followed by ANOVA
  - followed by other procedures or no post hoc
2. Use of ANOVA with sum scores to test group mean difference in multivariate data
3. Use of multiple ANOVAs to test group mean difference in multivariate data

Moreover, we provided a small simulation study to demonstrate the extent to which multiple ANOVAs yields incorrect results when they are inadvertently used to test group mean differences on multiple dependent variables. This simulation is also designed to determine to what extent the results of multiple ANOVAs agree to the results obtained from MANOVA. For the simulation conditions, data were generated from a standard multivariate normal distribution. Sample size is fixed to 100 for each group. Number of groups and number of dependent variables are fixed to two, and three, respectively. Correlation between the dependent variables, difference in the population means, and distribution variance are the three variables considered to create the simulation conditions. Correlation had two levels, which specifies lower- and higher-correlation conditions. More specifically, in lower correlation condition, the correlations between the dependent variables are drawn from a uniform distribution with minimum of .2 and maximum of .4. Likewise, correlations for the higher correlation condition are drawn from a uniform distribution with minimum value of .6 and maximum value of .8. Note that the mean of these distributions (i.e, .3 and .7) are the cutoff scores for describing the magnitude of a relationship in social sciences. As argued by Köklü, Büyüköztürk and Çokluk (2007), a correlation coefficient smaller than .3 represent a low relationship and one larger than .7 represents a high relationship.

Table 1. Variables Used in Simulation

Corr	$\Delta\mu$	$\sigma^2$
Lower = U(.2, .4)	Small = 0.2 standard deviation	Lower = .5
Higher = U(.6, .8)	Medium = 0.4 standard deviation	Medium = 1.0
	Large = 0.6 standard deviation	Higher = 1.5

Note: Corr is the correlations between the dependent variables;  $\Delta\mu$  is the population mean differences;  $\sigma^2$  is the distribution variance.

Population mean difference had three levels, which are labeled as small-, medium-, and large-difference conditions. These three levels were fixed to 0.2, 0.4, and 0.6 standard deviations. Here we have no intention to define what is a small or a large difference is; rather, we are just using these arbitrary differences to demonstrate the impact of the size of mean differences. More specifically, one

group is generated from a multivariate normal distribution  $MVN(0, \Sigma)$ , where  $\Sigma$  is the variance-covariance matrix determined by the variance of and correlations specified for each conditions. Then, 0.2, 0.4, or 0.6 is added to the mean vector of the second group for the small, medium, and large mean difference conditions, respectively. Last variable is the distribution variance for which 0.5, 1.0, and 1.5 were used to represent lower-, medium-, and higher-variance conditions. These three variables and their levels are summarized in Table 1. Combination of two correlation levels, three mean difference levels, and three variance levels yield 18 conditions. Number of replication for each condition is fixed to 500.

Table 2. The Extent to which Multiple ANOVAs are in Conformity with MANOVA

Corr	$\Delta\mu$	$\sigma^2$	MANOVA			ANOVAs		
			$p \geq .05$	$p \geq .05$	$p < .05$	$p < .05$	$p < .05$	$p \geq .05$
Lower	Small	Lower	.368	.174	.194	.632	.632	.000
		Medium	.676	.456	.220	.324	.322	.002
		Higher	.796	.558	.238	.204	.202	.002
	Medium	Lower	.004	.002	.002	.996	.996	.000
		Medium	.108	.026	.082	.892	.892	.000
		Higher	.258	.116	.142	.742	.742	.000
	Large	Lower	.000	.000	.000	1.000	1.000	.000
		Medium	.000	.000	.000	1.000	1.000	.000
		Higher	.024	.000	.024	.976	.976	.000
Higher	Small	Lower	.566	.322	.244	.434	.432	.002
		Medium	.798	.564	.234	.202	.196	.006
		Higher	.828	.626	.202	.172	.166	.006
	Medium	Lower	.032	.006	.026	.968	.968	.000
		Medium	.236	.056	.180	.764	.764	.000
		Higher	.446	.188	.258	.554	.554	.000
	Large	Lower	.000	.000	.000	1.000	1.000	.000
		Medium	.016	.002	.014	.984	.984	.000
		Higher	.096	.026	.070	.904	.904	.000

Note: Corr is the correlations between the dependent variables;  $\Delta\mu$  is the population mean differences;  $\sigma^2$  is the distribution variance; and  $p$  is the type I error rate of the test.

## RESULTS

### Simulation Results

Data generation and the analyses of the generated data are conducted in R language and statistical computing environment (R core team) using R-package "MASS" (Venables & Ripley, 2002). R code used for data generation and analyses is given in the Appendix A. Simulation results are summarized in Tables 2 and 3. These tables present the conformity on test results of MANOVA and multiple ANOVAs without and with Bonferroni correction, respectively. It should be noted here that, under the (multiple) ANOVAs condition, retain refers to the conditions where all three tests corresponding to three dependent variables are retained; whereas, reject refers to the conditions where at least one hypothesis out of the three is rejected. In the MANOVA tests, we used the Pillai's trace as rejection criterion because it is more robust to MANOVA violation of test assumptions (Olson, 1974).

First of all, result tables present two expected results: (1) Increase in the sample variance yields increase in the number of retained null hypotheses when the mean difference is tested by either multiple ANOVAs or by a MANOVA. For example, under the lower correlation and small mean difference cases, MANOVA retains about 37% to 80% of the null hypothesis as the variance increases from 0.5 to 1.5. Similarly, when we conduct multiple ANOVAs without Bonferroni correction, approximately 17% to 56% of the null hypotheses are retained as the sample variance increases from 0.5 to 1.5. Under the same conditions, when we conduct multiple ANOVAs with Bonferroni correction, these percentages become 34% (i.e., .318+.022) to 77% (i.e., .750+.016).

Table 3. The Extent to which Multiple Bonferroni Corrected ANOVAs Agree with MANOVA

Corr	$\Delta\mu$	$\sigma^2$	MANOVA	ANOVAs		MANOVA	ANOVAs	
			$p \geq .05$	$p \geq .0167$	$p < .0167$	$p < .05$	$p < .0167$	$p \geq .0167$
Lower	Small	Lower	.368	.318	.050	.632	.610	.022
		Medium	.676	.634	.042	.324	.290	.034
		Higher	.796	.750	.046	.204	.188	.016
	Medium	Lower	.004	.002	.002	.996	.996	.000
		Medium	.108	.088	.020	.892	.876	.016
		Higher	.258	.228	.030	.742	.710	.032
	Large	Lower	.000	.000	.000	1.000	1.000	.000
		Medium	.024	.000	.000	1.000	1.000	.000
		Higher	.024	.020	.004	.976	.976	.000
Higher	Small	Lower	.566	.458	.108	.434	.424	.010
		Medium	.798	.724	.074	.202	.180	.022
		Higher	.828	.766	.062	.172	.154	.018
	Medium	Lower	.032	.020	.012	.968	.966	.002
		Medium	.236	.152	.084	.764	.758	.006
		Higher	.446	.340	.106	.554	.548	.006
	Large	Lower	.000	.000	.000	1.000	1.000	.000
		Medium	.016	.008	.008	.984	.982	.002
		Higher	.096	.060	.036	.904	.902	.002

Note: Corr is the correlations between the dependent variables;  $\Delta\mu$  is the population mean differences;  $\sigma^2$  is the distribution variance; and p is the type I error rate of the test.

Another expected result is (2) the increase in the rejection rates of the tests along with the increase in the sample mean differences. For example, under the lower correlation and higher variance conditions, rejection rates of MANOVA varied from .204 to .976 as the sample mean differences increases from 0.2 standard deviation to 0.6 standard deviation. Rejection rates of multiple ANOVAs without Bonferroni correction vary between .440 (i.e., .238+.202) to 1.000 (i.e., .024+.976) for the same conditions. When ANOVAs are conducted with Bonferroni correction, rejection rates of multiple ANOVAs vary between .234 (i.e., .046+.188) to .980 (i.e., .004+.976). Although these are the expected results, we are more interested in the agreement between the MANOVA and multiple ANOVAs in terms of hypothesis test results. Remember that this simulation study only considers the similarity of the test results from a statistical point of view. We do not have any intention to downgrade the importance of theoretical considerations on choosing one or the other analysis.

When we look at the results obtained under lower and higher correlation conditions, MANOVA tend to fail to reject the null hypothesis as the correlation between the dependent variables increases. For example, when sample variance is higher and correlation between the dependent variables is lower, MANOVA retains the null hypothesis .796, .258, and .024 of the time for the small-, medium-, and large mean difference cases; whereas these rates rise up to .828, .446, and .096 under the higher correlation cases. As long as the simulation results concerned, we are mainly interested in the agreement rates of the two types of dependent variable analysis results. Looking at the retain rates, we observed a great quantity of disagreement under certain conditions. For instance, MANOVA retains the null hypotheses with a rate of .368 (i.e., 184 out of 500) under the lower correlation, small mean difference, and lower sample variance case. Multiple ANOVAs, however, only retain 87 out of the 184 null hypotheses, which are already retained by MANOVA (i.e., agreement on retaining the null hypotheses is .174). When Bonferroni correction is applied to ANOVA tests, this agreement rate is reported to be 159 out of 184 times (i.e., .318).

Tables 2 and 3 suggest that multiple ANOVAs procedure rejects a great deal of the null hypotheses that are already rejected by MANOVA. The highest disagreement rates for the ANOVAs are observed under small mean difference cases when Bonferroni correction is applied to ANOVAs (i.e., up to .034 and .022 under the lower and higher correlation conditions, respectively). In general, these results indicate that application of multiple ANOVAs rather than a single MANOVA yields higher rejection rates.

Results on Document Analysis

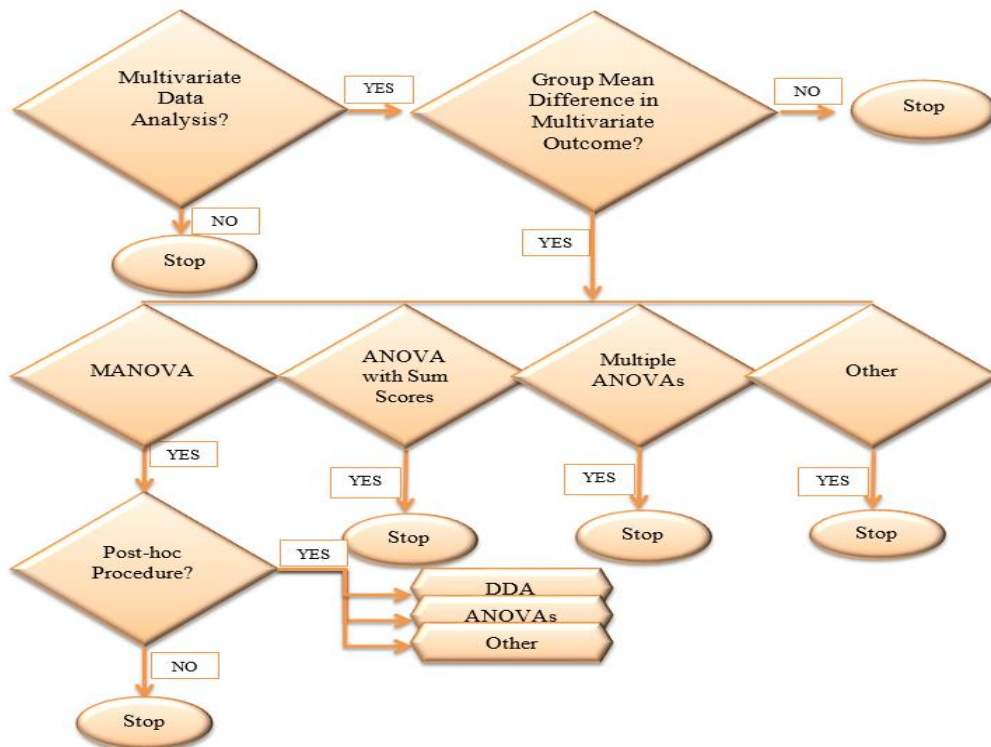


Figure 2. Flow Chart Used to Categorize the Reviewed Research.

We have gathered our archival data by screening the archives of three education journals (e.g., higher, medium, and lower quality). We have found 144 studies investigating the mean difference of multivariate dependent variables as we have viewed a total of 767 articles. We looked at the data analysis technique used for testing the group mean differences. To categorize reviewed works, we have used the flow chart given in Figure 2. In our archival survey, we have come across multiple t-tests applied to test the mean differences across two groups on multiple dependent variables. These studies were counted toward multiple ANOVAs category.

Table 4. Results on The Archival Survey

Journal Quality	Years	Number of Articles	Multivariate Mean Difference	MANOVA			Sum Score ANOVA	Multiple ANOVAs	
				No post hoc	ANOVA	DDA		No post hoc	Post hoc
HQ-J	2015	88	15	1	3	0	5	6	0
	2016	62	11	1	1	0	1	7	1
	2017	80	14	2	2	0	6	4	0
	2018	60	11	0	0	0	2	9	0
MQ-J	2015	61	13	0	2	0	4	7	0
	2016	50	12	1	2	0	2	7	0
	2017	60	3	0	0	0	0	3	0
	2018	45	10	0	2	0	4	4	0
LQ-J	2015	38	12	0	1	0	3	10	0
	2016	65	11	0	0	0	5	5	0
	2017	76	18	0	0	0	6	10	0
	2018	82	14	0	1	0	4	9	0
All-3-J	15-18	767	144	5	14	0	42	81	1

Note: HQ-J = higher quality journal; MQ-J = medium quality journal; LQ-J = lower quality journal; All-3-J = all three journals.

Results on the archival survey are summarized in Table 4. Rate of articles investigating treatment variable effect on the multivariate variables are about 18% (i.e., 51/290 and 38/216) for the higher and medium quality journals, while this rate is slightly higher for the lower quality journal (i.e., 21% or 55/261). Rate of MANOVA test use for detecting treatment effect is quite low: 10/51; 7/38; and 2/55 for the higher, medium, and lower quality journal publications, respectively. Although the maximum number of studies investigating mean differences on multivariate data is reported to be published in the lower quality journal, use of MANOVA to test the mean difference is only about 4% (i.e., 2 out of 55). Within the rare use of MANOVA, employment of ANOVA as post hoc tests is quite common (i.e., 14 out of 19). This may be mainly due to the fact that ANOVA tests are readily available when MANOVA test is run by the statistical package for the social sciences (SPSS). Moreover, although the most accurate inferences can be made when DDA is run as a follow up test for MANOVA, we have not come across any study that used DDA to interpret MANOVA results.

It is obvious from the results summarized in Table 4 that many researchers do not use MANOVA when it is the most appropriate way to test effect of independent variable(s) on the multivariate dependent variables. Rather than using MANOVA, many educational researchers who published in Turkish educational journals run a single ANOVA on the sum score obtained from multivariate dependent variables or they run multiple ANOVAs to test the effect on each of the dependent variables separately. Figure 3 displays these results based on the three types of journals as well as the results obtained from all three journals altogether. This figure shows that employment of MANOVA is quite rare across all, especially for the lower, quality journal publications. At least more than half of the studies run multiple ANOVAs rather than running a single MANOVA to test group mean differences on the multivariate dependent variables. Furthermore, approximately 30% of the studies used a single ANOVA test on a dependent variable, which is obtained by summing all the scores on multiple dependent variables.

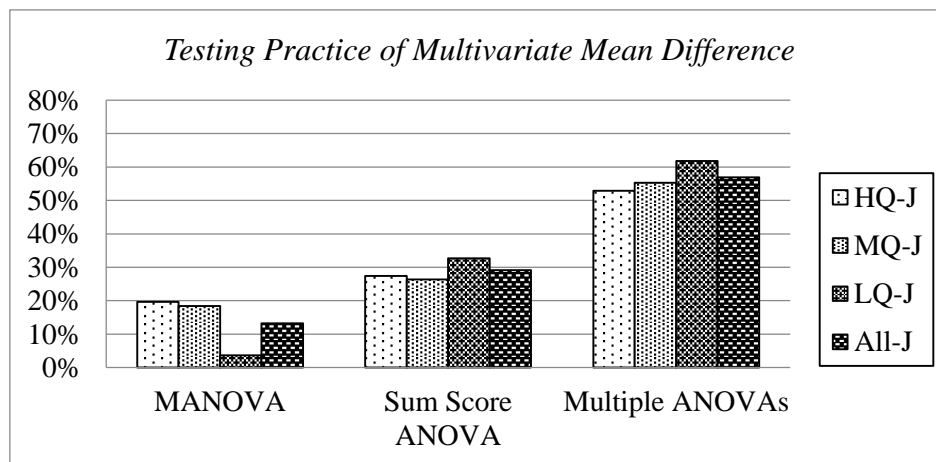


Figure 3. Rate of Analyses Used to Test Multivariate Mean Differences.

## DISCUSSION and CONCLUSION

Primary purpose of conducting univariate or multivariate analysis of variance is to determine treatment variable effects. Although MANOVA can be considered as a more general procedure of ANOVA, it is not just a statistical extension of ANOVA because they address different research questions. ANOVA is used to test the group mean differences on an observed variable whereas MANOVA is used to test the group difference on an underlying latent variables. By conducting a MANOVA we basically test the group mean differences on a linear combination of the dependent variables. Because we are not interested in the mean difference of any single dependent variable when we conduct MANOVA, conducting multiple ANOVAs (i.e., an ANOVA for each dependent variable) would not



be the same as conducting a single MANOVA. To do so would not address the empirical questions researchers begins with and yield different statistical test results.

With this study, we aimed to emphasize the theory behind the MANOVA and to make clear distinction between surrogate statistical procedures such as ANOVA. We not only focused on the theoretical difference between the two; through a small simulation study, we also demonstrated the discrepancy between obtained statistical test results. Then, we further investigated the extent of incorrect analysis of multivariate data in educational studies that are published in Turkish education journals. We specifically focused on the analysis of multivariate data for treatment variable effects and the post hoc procedures used for follow up. Results indicated that correct use of MANOVA with its proper post hoc procedure is not common practice across educational researchers who publish in Turkish education journals.

Although the courses given in the graduate level include the analysis of multivariate data, it is observed that, at least in case of MANOVA, the areas of application are not properly understood. The underlying reason for this may be the presentation of practical information on how to analyze data at hand with specific statistical package programs (eg., SPSS) rather than presentation of the theoretical background of these statistical data analysis techniques. In order to eliminate such deficiencies and misunderstandings of individuals who are conducting research in education, it is useful to take steps to gain theoretical knowledge on the basis of statistical analysis in the graduate education programs. We also suggest researchers to co-operate with the experts of the related fields if they deem necessary.

## REFERENCES

- Bowen, G. A., (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. London; Routledge.
- Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development*, 21, 130-137.
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analysis. *Psychological Bulletin*, 105, 302-308.
- Hummel, T. J., & Sligo, J. R. (1971). Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 76, 49-57.
- Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *Journal of Experimental Education*, 69, 280-309.
- Köklü, N., Büyükoztürk, Ş., & Çokluk, Ö. (2007). *Sosyal bilimler için istatistik*. Ankara: Pegem Yayınları.
- Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of The American Statistical Association*, 69(348), 894-908.
- R Core Team (2013). R: A language and environment for statistical computing [Computer software]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Sherry, A. (2006). Discriminant analysis in counseling psychology research. *The Counseling Psychologist*, 34, 661-683. doi:10.1177/0011000006287103
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tonidandel, S., & LeBreton, J. M. (2013). Beyond step-down analysis: A new test for decomposing the importance of dependent variables in MANOVA. *Journal of Applied Psychology*, 98, 469-477.
- Venables, W. N. & Ripley, B. D. (2002) *Modern applied statistics with S* (Fourth Edition). New York, NY: Springer.
- Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Assessment, Research & Evaluation*, 19(17), 1-10.
- Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, 56, 134-149.
- Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, 38, 343-352.

## ANOVA'nın Eğitim Araştırmalarında Dikkatsizce Kullanımı: ANOVA, MANOVA için Yer Tutucu Değildir

### Giriş

Tek değişkenli varyans analizi (ANOVA) ve çok değişkenli varyans analizi (MANOVA) farklı araştırma sorularına cevap arayan iki farklı istatistiksel yöntemdir. Bu ikisi arasındaki seçim araştırmanın amacına bağlı olarak yapılır; tek bir bağımlı değişken için gruplar arası fark bakılırken ANOVA, birden fazla bağımlı değişken için gruplar arası fark bakılırken MANOVA'dan bahsediyoruzdur (Fish, 1988; Stevens, 2002). MANOVA istatistiksel olarak ANOVA'nın daha genel bir prosedürü olarak düşünülebilir. MANOVA en sık kullanılan çok değişkenli data analiz prosedürlerinden biri olsa da (Kieffer, Reese & Thompson, 2001; Zientek & Thompson, 2009); alan yazın incelendiğinde bu prosedür ve analize eşlik etmesi gereken doğru post hoc prosedürünün azımsanmayacak sayıda sosyal bilimler araştırmacısı tarafından doğru anlaşılmadığı görülmektedir (Tonidandel & LeBreton, 2013; Warne, 2014; Warne, Lazo, Ramos & Ritter, 2012).

MANOVA testinin kullanılması gereken yerlerde en sık karşımıza çıkan yanlış kullanım her bir bağımlı değişkeni ayrı ayrı test eden ANOVA testleri serisinin tercih edilmesidir. Ancak, birden fazla ANOVA testinin uygulanması bir tek MANOVA testinin uygulanmasından iki şekilde farklılık arz eder: (1) birden fazla ANOVA uygulaması *birinci tip hatasının* yapılma olasılığını artırır. Bu hatanın artış oranı bağımlı değişkenler arasındaki korelasyonun büyüklüğü ile değişmekte olup kolayca kontrol altına alınamaz. Tamamen bağımsız yani korelasyonun sıfır olduğu durumlar için Bonferroni düzeltmesi uygulamak bu hata oranının ancak kontrol altına alınmasını sağlayabilir (Hummel & Sligo, 1971) ki sosyal bilimlerdeki çoklu bağımsız değişkenler arasındaki korelasyonun sıfır olduğu durum (eğer varsa) sınırlıdır.

Çoklu ANOVA ve MANOVA arasındaki ikinci temel fark ise (2) bu testlerin farklı ampirik sorulara cevap verebilir olmasıyla ilgilidir. ANOVA gözlenen değişkenlerden elde edilen veriler için uygun bir test iken; MANOVA gözlenmeyen (gizil) değişkenler üzerinden gruplar arası farklılık olup olmadığını anlamak için yapılabilecek uygun bir testtir (Zientek & Thompson, 2009). Birden fazla ANOVA testinden elde edilen sonuçlar bağımsız değişken(ler) ile bağımlı değişkenlerin kombinasyonu arasında anlamlı bir ilişki olup olmadığını test etmede yetersiz kalır (Warne, 2014). MANOVA testinin kullanımında araştırmacılar gözlenmeyen değişkenlerin gözlenen gösterge (indicator) değişkenleri açısından gruplar arasında fark olup olmadığını değil, bu gösterge değişkenlerin lineer bir kombinasyonundan oluşan gözlenemeyen değişken açısından gruplar arasında anlamlı bir farklılık olup olmadığını araştırmaktadır.

MANOVA yerine yanlışlıkla ANOVA kullanımının bir diğer şekli ise bağımlı değişkenlerden elde edilen skorların toplamı üzerinden bir tek ANOVA testinin yapılmasıdır. Bu çalışmanın iki temel amacı vardır. (1) MANOVA'nın ve devamında uygulanması gereken post hoc testinin alt yapısını oluşturan teoriyi vurgulayarak ANOVA ve MANOVA arasındaki farklılıkların anlaşılmasına yardımcı olmak; (2) Türkiye'de yayınlanan eğitim dergilerinde basılmış makalelerde MANOVA testinin ve doğru post hoc testinin kullanılması gerektiği durumlarda bunların kullanılmış olma oranını ortaya koymaktır.

### Yöntem

Çalışmanın yöntemi betimsel araştırma yöntemlerinden doküman analizidir. Doküman analizi içerik analizine yakın bir veri analizi yöntemidir. Bu yöntem basılı ya da elektronik materyallerin sistematik bir şekilde incelenmesinin ve değerlendirilmesinin yapılması şeklinde tanımlanabilir (Bowen, 2009). Bu doküman analiziyle araştırmacılar, Türkiye'de yayın yapan eğitim dergilerinde basılmış makalelerde, MANOVA testinin ne ölçüde doğru kullanıldığının tespitini yapmayı amaçlamaktadırlar. Üç eğitim dergisinin arşivlerinden son dört yılda (2015-2018) yayınlanan tüm sayıları incelemek

kaydıyla veriler elde edilmiştir. Bu üç dergi, endekslandıkları yerlere göre yüksek, orta ve düşük kaliteli dergileri temsil edecek şekilde seçilmiştir. Bu sınıflandırma dergilerin tarandıkları veri tabanları göz önünde bulundurularak yapılmıştır (ör. Thompson Reuters Sosyal Bilimler Atıf Dizini yüksek kaliteli dergiyi endekslemektedir). Araştırmacılar bu üç Türk eğitim dergisinde yayınlanan makaleleri incelemiş çok değişkenli bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişkiye bağlı olarak gruplar arası anlamlı farklılıkları test etmek için kullanılan analiz türlerini raporlaştırmışlardır.

Tablo 1. Simulasyonda Manipüle Edilen Değişkenler

Corr	$\Delta\mu$	$\sigma^2$
Düşük = U(.2, .4) Yüksek = U(.6, .8)	Küçük = 0.2 standart sapma Orta = 0.4 standart sapma Büyük = 0.6 standart sapma	Düşük = .5 Orta = 1.0 Yüksek = 1.5

Not: Corr, bağımlı değişkenler arasındaki korelasyon;  $\Delta\mu$ , popülasyon ortalamalarındaki fark;  $\sigma^2$ , dağılımların varyansı.

Ayrıca, çoklu ANOVA'ların yanlışlıkla çok değişkenli bağımlı değişkenler üzerindeki grup ortalama farklarını test etmek için kullanıldığında istatistiksel olarak ne ölçüde tutarlı sonuçlar verdiğini gösteren küçük bir simülasyon çalışması yaptık. Simülasyon koşulları için, standart çok değişkenli normal dağılımdan veriler üretilmiştir. Örneklem büyüklüğü her grup için 100'e sabitlenmiştir. Grup sayısı ve bağımlı değişken sayısı sırasıyla iki ve üçe sabitlenmiştir. Bağımlı değişkenler arasındaki korelasyon, popülasyon ortalamaları arasındaki fark ve dağılım varyansı, simülasyon koşullarını oluşturmak için manipüle edilen değişkenlerdir. Bu üç değişken ve değişkenlerin düzeyleri Tablo 1'de özetlenmiştir. İki korelasyon düzeyi, üç ortalama fark düzeyi ve üç varyans düzeyinin çaprazlanmasıyla toplam 18 simülasyon durumu oluşturulmuştur. Her durum için replikasyon sayısı 500 olarak belirlenmiştir.

### Sonuç ve Tartışma

Düşük ve yüksek korelasyon koşulları altında elde edilen simülasyon sonuçlarına baktığımızda, MANOVA bağımlı değişkenler arasındaki korelasyon arttıkça yokluk hipotezini daha sıklıkla reddetme eğilimindedir. Bu simülasyon sonuçları içinden biz bağımlı değişken analizinde kullanılan iki tür testin (MANOVA ve Çoklu ANOVA) sonucunun mutabakat oranlarıyla daha çok ilgileniyoruz. Mutabakat oranlarına bakıldığında, belirli koşullar altında büyük miktarda anlaşmazlık olduğunu gözlemleyebiliriz. Örneğin, MANOVA yokluk hipotezini düşük korelasyon, küçük popülasyon ortalama farkı ve düşük dağılım varyansı durumunda .368 oranında reddedemektedir. Bununla birlikte, çoklu ANOVA'lar, reddedilemeyen yokluk hipotezlerinin en az yarısını reddetmektedir. Bonferroni düzeltmesi ANOVA testlerine uygulandığında, MANOVA ve çoklu ANOVA arasındaki yokluk hipotezlerini reddedememe mutabakatlarının oldukça yükseldiği gözlenmiştir. Simülasyon sonuçları çoklu ANOVA ve MANOVA'nın yokluk hipotezini reddetme mutabakatlarının oldukça yüksek olduğu sonucunu ortaya koymaktadır. Birkaç istisna dışında, üç ANOVA'dan en az biri, MANOVA tarafından zaten reddedilmiş yokluk hipotezlerini reddetmektedir. Genel olarak, bu sonuçlar tek bir MANOVA yerine birden fazla ANOVA uygulamasının daha yüksek oranda yokluk hipotezi reddetme eğilimi gösterdiğini ortaya koymaktadır.

Üç eğitim dergisinin arşivleri taranarak çok değişkenli bağımlı değişkenlerin grup ortalama farkını araştıran 144 çalışma bulunmuştur. Çok değişkenli bağımlı değişkenler üzerinde bağımsız değişkeninin etkisini araştıran makalelerin oranı, yüksek ve orta kaliteli dergiler için yaklaşık %18 (yani, 51/290 ve 38/216) iken, düşük kaliteli dergi için %21 (55/261) olarak bulunmuştur. Bağımsız değişken etkisinin saptanmasında MANOVA testi kullanım oranının oldukça düşük olduğu görülmüştür: 10/51; 7/38; ve 2/55 sırasıyla yüksek, orta ve düşük kaliteli dergiler için. MANOVA'nın nadir kullanımı içinde, ANOVA'nın post hoc testi olarak kullanımının oldukça yaygın olduğu görülmüştür (14/19). Bu durum MANOVA testinin sosyal bilimler için istatistiksel paket (SPSS) programı tarafından gerçekleştirildiğinde, ANOVA testlerinin otomatik olarak uygulanıyor

olmasından kaynaklanıyor olabilir. Bununla birlikte, DDA MANOVA için en doğru post hoc prosedürü olmasına rağmen, MANOVA sonuçlarını yorumlayabilmek için post hoc olarak DDA kullanılan herhangi bir çalışmaya rastlanmamıştır.

MANOVA'yı kullanmak yerine, Türk eğitim dergilerinde yayınlanan birçok eğitim araştırmacısı, çok değişkenli bağımlı değişkenlerden elde edilen toplam puan üzerinde tek bir ANOVA testini uygulamakta veya bağımlı değişkenlerin her biri üzerindeki bağımsız değişken etkisini ayrı ayrı test etmek için birden fazla ANOVA testi kullanmaktadır. Sonuçlar, MANOVA'nın uygulamasının bütün dergi türlerinde oldukça nadir olduğunu göstermektedir. Çalışmaların yarısından fazlası, çok değişkenli bağımlı değişkenlerdeki grup ortalama farklarını test etmek için tek bir MANOVA çalıştırmak yerine birden fazla ANOVA kullanıyor. Ayrıca, çalışmaların yaklaşık %30'u, çoklu bağımlı değişkenlerden elde edilen toplam puanlar üzerinden tek bir ANOVA testi yaparak bağımsız değişkenlerin etkisini ortaya çıkarmaya çalışmaktadır. Bütün bu sonuçlar bize MANOVA'nın teorisinin ve uygulamasının ülkemizdeki eğitim dergilerinde yayın yapan eğitim araştırmacılarınca yeterince anlaşılmadığını göstermektedir.

Lisansüstü eğitim dönemlerinde her ne kadar çok değişkenli verilerin analizini içeren dersler veriliyor olsa da MANOVA açısından bakıldığında, en azından uygulama alanlarının yeterince iyi anlaşılmadığı görülmektedir. Bunun altında yatan temel sebep, istatistiksel veri analizi yöntemlerinin teorik alt yapısından ziyade, belirli istatistiksel paket programlar (ör. SPSS) ile nasıl analiz yapılacağına ilişkin pratik bilgilerin sunuluyor olması olabilir. Eğitimde araştırma yapan bireylerin bu tür eksik ve yanlışlarının giderilmesi için lisansüstü eğitim programlarının istatistiksel analizlerin dayandığı teorik bilgileri kazandırmaya yönelik adımlar atması ve eğitim araştırmacılarının da gerekli gördükleri durumlarda ilgili alanların uzmanlarıyla iş birliğine yönelmeleri faydalı olabilir.

## Appendix A. R Code Used for Data Generation and Analyses

```

library(MASS)
d=3 ### dimensions...fixed
N=100 ##### sample size...fixed
M_fark=.2 ##### group mean differences...variable
variance=.5 ##### distribution variance...variable
mincorr=.2 ### minimum correlation...variable
maxcorr=.4 ### maximum correlation...variable
var=matrix(variance,d)
std=sqrt(var)
R=500 ### number of replication...fixed

Pvalues<-matrix(NA,R,d+1)
for (r in 1:R){
  corr=matrix(runif(d,mincorr,maxcorr))
  corMat=matrix(c(1,corr[1,],corr[2,],corr[1,],1,corr[3,],corr[2,],corr[3,],1),ncol=d)
  covMat=std%*%t(std)*corMat
  Mean1=matrix(0,d)
  Mean2=Mean1+M_fark
  data1=mvnrm(N,Mean1,covMat)
  data2=mvnrm(N,Mean2,covMat)
  data=rbind(data1,data2)
  ind=rbind(matrix(1,N),matrix(2,N))
  ##### MANOVA
  fit <-summary(manova(data ~ ind) , test="Pillai")
  p<-fit$stats[1:1, "Pr(>F)"]
  p<- matrix(unlist(p), nrow=length(p))
  ##### ANOVAs
  values1=data[,1]
  values2=data[,2]
  values3=data[,3]
  aov_sum1=summary(aov(values1~ind))
  p1=lapply(aov_sum1, function(aov_sum1){aov_sum1$'Pr(>F)' })
  p1<- matrix( unlist(p1), nrow=length(p1) )
  aov_sum2=summary(aov(values2~ind))
  p2=lapply(aov_sum2, function(aov_sum2){aov_sum2$'Pr(>F)' })
  p2<- matrix( unlist(p2), nrow=length(p2) )
  aov_sum3=summary(aov(values3~ind))
  p3=lapply(aov_sum3, function(aov_sum3){aov_sum3$'Pr(>F)' })
  p3<- matrix( unlist(p3), nrow=length(p3) )
  pval=round(cbind(p[,1],p1[,1],p2[,1],p3[,1]), digits=3) ### manova + 3 anovas
  Pvalues[r,]<-pval
}
ret<-ifelse(Pvalues>=.05,1,0) ##### Retained hypotheses rates of MANOVA and ANOVAs
manret<-ifelse(ret[,1]==1,3,9)
anoret<-ifelse(rowSums(ret[,2:4])==3,3,99)
agree<-as.numeric(sum(ifelse(manret==anoret,1,0)))
disagree<-as.numeric((sum(ret[,1])-agree))
match<-as.matrix(cbind(agree,disagree))
rej<-ifelse(Pvalues<.05,1,0)
anorej<-ifelse(rowSums(rej[,2:4])>0,1,9)
agreement<-as.numeric(sum(ifelse(anorej==rej[,1],1,0))) ### Not manova but at least one anova is significant
disagreement<-as.numeric((sum(rej[,1])-agreement))
matching<-as.matrix(cbind(agreement,disagreement)) ##### retained manova while rejected anova
results<-cbind(match,matching)

#Bonferroni corrected
retB<-ifelse(Pvalues>=.0167,1,0) ##### Retained hypotheses rates of MANOVA and ANOVAs
anoretB<-ifelse(rowSums(retB[,2:4])==3,3,99)
agreeB<-as.numeric(sum(ifelse(manretB==anoretB,1,0)))
disagreeB<-as.numeric((sum(retB[,1])-agreeB))
matchB<-as.matrix(cbind(agreeB,disagreeB))
rejB<-ifelse(Pvalues<.0167,1,0)
anorejB<-ifelse(rowSums(rejB[,2:4])>0,1,9)
agreementB<-as.numeric(sum(ifelse(anorejB==rejB[,1],1,0))) ### Not manova but at least one anova is significant
disagreementB<-as.numeric((sum(rejB[,1])-agreementB))
matchingB<-as.matrix(cbind(agreementB,disagreementB)) ##### retained manova while rejected anova
resultsB<-cbind(matchB,matchingB)
cbind(results,resultsB)

```