2004

Multivariate Analysis Of Variance And Covariance (Manova And Mancova)

Gudmund R. Iversen
Swarthmore College, iversen@swarthmore.edu

Recommended Citation
Gudmund R. Iversen. (2004). "Multivariate Analysis Of Variance And Covariance (Manova And Mancova)". *The SAGE Encyclopedia Of Social Science Research Methods.* Volume 2, 702-703. DOI: 10.4135/9781412950589.n608
https://works.swarthmore.edu/fac-math-stat/212

SAGE reference

The SAGE Encyclopedia of Social Science Research Methods

MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE (MANOVA AND MANCOVA)

Contributors: Gudmund R. Iversen
Edited by: Michael S. Lewis-Beck, Alan Bryman & Tim Futing Liao
Book Title: The SAGE Encyclopedia of Social Science Research Methods
Chapter Title: "MULTIVARIATE ANALYSIS OF VARIANCE AND COVARIANCE (MANOVA AND MANCOVA)"
Pub. Date: 2004
Access Date: March 13, 2020
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780761923633
Online ISBN: 9781412950589
DOI: http://dx.doi.org/10.4135/9781412950589.n608
Print page: 703

The word *multi* here refers to the use of several DEPENDENT VARIABLES in a single analysis. This is in spite of the often-used term for several EXPLANATORY variables, as in MULTIPLE REGRESSION. These multiple analyses can successfully be used in fields as separate as psychology and biology—in psychological testing, each test item becomes a separate dependent variable, and in biology, there are several observations on a particular organism.

MANOVA and MANCOVA are models for the joint statistical analysis of several QUANTITATIVE dependent variables in one analysis, using the same explanatory variables for all the dependent variables. In MANOVA, all the explanatory variables are nominal variables, whereas in MANCOVA, some of the explanatory variables are quantitative and some are qualitative (nominal). These models can also be extended to the regression case in which all the explanatory variables are quantitative. These three approaches are special cases of the so-called multivariate general linear model.

With several dependent variables and a set of explanatory variables, it is possible to separately analyze the relationship of each dependent variable to the explanatory variables, using ANALYSIS OF VARIANCE or ANALYSIS OF COVARIANCE. The difference in MANOVA and MANCOVA is that here all the dependent variables are used in the same, one analysis. This can be desirable when the several dependent variables have something in common, such as scores on different items in a psychological test. This multivariate approach also eliminates the problem of what to do for an overall significance level when many statistical tests are run on the same data.

Explanation

As a simple example, suppose there are data on both the mathematics and verbal Scholastic Aptitude Test (SAT) scores for a group of students. Are there gender and race differences in the two sets of SAT scores? It is possible to do a separate ANALYSIS OF VARIANCE for each SAT score, as expressed in the following two equations:

$$\begin{aligned}\text{Math SAT} = {} & \text{Gender} + \text{Race} + \text{Gender} \cdot \text{Race} \\ & + \text{Residual},\end{aligned}$$

$$\begin{aligned}\text{Verbal SAT} = {} & \text{Gender} + \text{Race} + \text{Gender} \cdot \text{Race} \\ & + \text{Residual}.\end{aligned}$$

Each analysis would provide the proper SUMS OF SQUARES and *P* VALUES, and it would be possible to conclude what the impacts are on the SAT scores of the two explanatory variables and their STATISTICAL INTERACTIONS.

However, two such separate analyses do not use the fact that the two explanatory variables themselves are related. If there were some way of taking this information into account, it would be possible to perform a stronger analysis and perhaps even find significant results that were not apparent in the two separate analyses. This can now be done using *one* MANOVA analysis instead of *two* separate ANOVAs.

Historical Account

The computational complexity increases dramatically by going from analysis of variance and analysis of covariance to the corresponding multivariate analyses. Thus, the methods were not well developed and used until the necessary statistical software became available in the latter parts of the past century. Now, all major statistical software packages have the capacity to perform MANOVA and MANCOVA. Also, many of the software manuals have good explanations of the methods and how to interpret the outputs.

Applications

Let us consider the two SAT scores and one explanatory variable, say, gender. With two separate analyses of variance, we study the distributions of the math and verbal scores separately, and we may or may not find gender differences. However, with two dependent variables, we can now create a scatterplot for the two test scores, showing their relationship. To identify the gender of each respondent, we can color the points in two different colors. It will then be clear whether the points for the two groups overlap, or the groups may show up as two very different sets of points in a variety of possible ways. We can now see differences between the two groups that were not obvious when each variable was only considered separately. In the extreme, it is even possible to have the same mean value for both females and males on one of the two test scores. However, in the scatterplot, we may see the way in which the two groups of scores differ in major ways. Such differences would be found using one multivariate analysis of variance instead of two separate analyses.

The first step in both MANOVA and MANCOVA is to test the overall NULL HYPOTHESIS that all groups have the same means on the various dependent variables. In the example above, that implies testing that females and males have the same math scores and verbal scores and that all races have the same math scores and verbal scores. If these null hypotheses are rejected, the next step is to find *which* group means are different, just as we do in the univariate case.

Typically, there are four different ways to test the overall null hypothesis. They are known as Wilks's lambda, the Pillai-Bartlett trace, Roy's greatest characteristic root, and the Hotelling-Lawley trace. The four methods may not agree in their test of the same null hypotheses, and the choice is usually governed by issues having to do with robustness of the tests and their statistical power. Wilks's test is the most commonly used, and his test statistic has a distribution that can be approximated by an F distribution when the proper assumptions are met.

Assumptions

As with other statistical procedures, the data have to satisfy certain assumptions for the procedure to work, especially to produce meaningful p values. The data need to form a proper statistical random sample from an underlying population. The observations need to be obtained independently of one another. The dependent variables need to have a multivariate normal distribution, which may be the case when each dependent variable has a normal distribution. Finally, homogeneity of variance must be the case for each dependent variable, and the correlation between any two dependent variables must be the same in all groups of observations.

AUTHOR'S NOTE: A search on the World Wide Web of MANOVA will bring up several good explanations and examples of uses of these methods.

- multiple analysis of variance
- dependent variables
- analysis of variance
- null hypothesis
- variance and covariance
- analysis of covariance
- p value

Gudmund R. Iversen
http://dx.doi.org/10.4135/9781412950589.n608
10.4135/9781412950589.n608

References

Bray, J. H., & Maxwell, S. E.(1985). Multivariate analysis of variance (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–054). Beverly Hills, CA: Sage.
Velleman, P. F.(1988). DataDesk version 6.0 statistics guide. Ithaca, NY: Data Description, Inc.