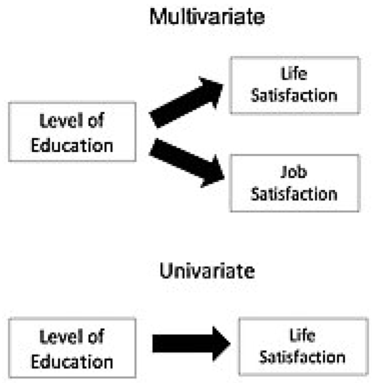
Multivariate analysis of variance

In statistics, **multivariate analysis of variance** (**MANOVA**) is a procedure for comparing multivariate sample means. As a multivariate procedure, it is used when there are two or more dependent variables,[1] and is often followed by significance tests involving individual dependent variables separately.[2]



The image above depicts a visual comparison between multivariate analysis of variance (MANOVA) and univariate analysis of variance (ANOVA). In MANOVA, researchers are examining the group differences of a singular independent variable across multiple outcome variables, whereas in an ANOVA, researchers are examining the group differences of sometimes multiple independent variables on a singular outcome variable. In the provided example, the levels of the IV might include high school, college, and graduate school. The results of a MANOVA can tell us whether an individual who completed graduate school showed higher life AND job satisfaction than an individual who completed only high school or college. Results of an ANOVA can only tell us this information for life satisfaction. Analyzing group differences across multiple outcome variables often provides more accurate information as a pure relationship between only X and only Y rarely exists in nature.

Contents

Relationship with ANOVA

MANOVA is a generalized form of univariate analysis of variance (ANOVA),[1] although, unlike univariate ANOVA, it uses the covariance between outcome variables in testing the statistical significance of the mean differences.

Where sums of squares appear in univariate analysis of variance, in multivariate analysis of variance certain positive-definite matrices appear. The diagonal entries are the same kinds of sums of squares that appear in univariate ANOVA. The off-diagonal entries are corresponding sums of products. Under normality assumptions about error distributions, the counterpart of the sum of squares due to error has a Wishart distribution.

MANOVA is based on the product of model variance matrix, Σ_{model} and inverse of the error variance matrix, Σ_{res}^{-1}, or $A = \Sigma_{\text{model}} \times \Sigma_{\text{res}}^{-1}$. The hypothesis that $\Sigma_{\text{model}} = \Sigma_{\text{residual}}$ implies that the product $A \sim I$.[3] Invariance considerations imply the MANOVA statistic should be a measure of magnitude of the singular value decomposition of this matrix product, but there is no unique choice owing to the multi-dimensional nature of the alternative hypothesis.

The most common[4][5] statistics are summaries based on the roots (or eigenvalues) λ_p of the A matrix:

- Samuel Stanley Wilks' $\Lambda_{\text{Wilks}} = \prod_{1,\ldots,p}(1/(1+\lambda_p)) = \det(I+A)^{-1} = \det(\Sigma_{\text{res}})/\det(\Sigma_{\text{res}}+\Sigma_{\text{model}})$ distributed as lambda (Λ)
- the K. C. Sreedharan Pillai–M. S. Bartlett trace, $\Lambda_{\text{Pillai}} = \sum_{1,\ldots,p}(\lambda_p/(1+\lambda_p)) = \operatorname{tr}(A(I+A)^{-1})$[6]
- the Lawley–Hotelling trace, $\Lambda_{\text{LH}} = \sum_{1,\ldots,p}(\lambda_p) = \operatorname{tr}(A)$
- Roy's greatest root (also called *Roy's largest root*), $\Lambda_{\text{Roy}} = \max_p(\lambda_p)$

Discussion continues over the merits of each,[1] although the greatest root leads only to a bound on significance which is not generally of practical interest. A further complication is that, except for the Roy's greatest root, the distribution of these statistics under the null hypothesis is not straightforward and can only be approximated except in a few low-dimensional cases.[7] An algorithm for the distribution of the Roy's largest root under the null hypothesis was derived in [8] while the distribution under the alternative is studied in.[9]

The best-known approximation for Wilks' lambda was derived by C. R. Rao.

In the case of two groups, all the statistics are equivalent and the test reduces to Hotelling's T-square.

Correlation of dependent variables

MANOVA's power is affected by the correlations of the dependent variables and by the effect sizes associated with those variables. For example, when there are two groups and two dependent variables, MANOVA's power is lowest when the correlation equals the ratio of the smaller to the larger standardized effect size.[10]

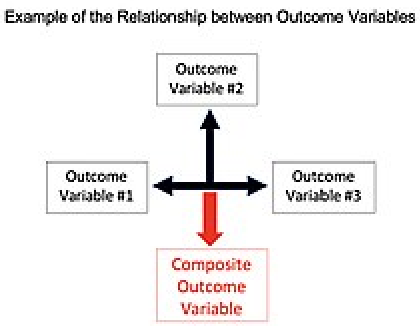
See also

- Discriminant function analysis
- Canonical correlation analysis
- Multivariate analysis of variance (Wikiversity)
- Repeated measures design

References

1. Warne, R. T. (2014). "A primer on multivariate analysis of variance (MANOVA) for behavioral scientists" (http://pareonline.net/getvn.asp?v=19&n=17). *Practical Assessment, Research & Evaluation*. **19** (17): 1–10.
2. Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erblaum.
3. Carey, Gregory. "Multivariate Analysis of Variance (MANOVA): I. Theory" (http://ibgwww.colorado.edu/~carey/p7291dir/handouts/manova1.pdf) (PDF). Retrieved 2011-03-22.
4. Garson, G. David. "Multivariate GLM, MANOVA, and MANCOVA" (http://faculty.chass.ncsu.edu/garson/PA765/manova.htm). Retrieved 2011-03-22.

5. UCLA: Academic Technology Services, Statistical Consulting Group. "Stata Annotated Output – MANOVA" (http://www.ats.ucla.edu/stat/stata/output/Stata_MANOVA.htm). Retrieved 2011-03-22.
6. "MANOVA Basic Concepts – Real Statistics Using Excel" (http://www.real-statistics.com/multivariate-statistics/multivariate-analysis-of-variance-manova/manova-basic-concepts/). *www.real-statistics.com*. Retrieved 5 April 2018.
7. Camo http://www.camo.com/multivariate_analysis.html
8. Chiani, M. (2016), "Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance", *Journal of Multivariate Analysis*, **143**: 467–471, arXiv:1401.3987v3 (https://arxiv.org/abs/1401.3987v3), doi:10.1016/j.jmva.2015.10.007 (https://doi.org/10.1016%2Fj.jmva.2015.10.007)
9. I.M. Johnstone, B. Nadler "Roy's largest root test under rank-one alternatives" arXiv preprint arXiv:1310.6581 (2013)
10. Frane, Andrew (2015). "Power and Type I Error Control for Univariate Comparisons in Multivariate Two-Group Designs". *Multivariate Behavioral Research*. **50** (2): 233–247. doi:10.1080/00273171.2014.968836 (https://doi.org/10.1080%2F00273171.2014.968836). PMID 26609880 (https://pubmed.ncbi.nlm.nih.gov/26609880).



This is a graphical depiction of the required relationship amongst outcome variables in a multivariate analysis of variance. Part of the analysis involves creating a composite variable, which the group differences of the independent variable are analyzed against. The composite variables, as there can be multiple, are different combinations of the outcome variables. The analysis then determines which combination shows the greatest group differences for the independent variable. A descriptive discriminant analysis is then used as a post hoc test to determine what the makeup of that composite variable is that creates the greatest group differences.

External links

- Multivariate Analysis of Variance (MANOVA) by Aaron French, Marcelo Macedo, John Poulsen, Tyler Waterson and Angela Yu, San Francisco State University (http://online.sfsu.edu/~efc/classes/biol710/manova/manovanewest.htm)
- What is a MANOVA test used for? (https://spss-tutor.com/manova.php)