Multivariate Analysis of Variance

Multivariate analysis of variance (MANOVA) compares groups on a set of dependent variables simultaneously. Rather than test group differences using several separate ANOVAs and run the risk of increased familywise error (probability of one or more Type I errors), the MANOVA approach makes a single comparison.[1] MANOVA is appropriate only when the several dependent variables are related to one another and the pattern of group differences expected for all of the dependent variables is in the same direction. The multiple measures can be several scale scores, individual items, or other related measures. An example might be a researcher's interest in which of several psychotherapy approaches (independent variable) differ in their ability to reduce psychological distress, where several measures of psychological distress, including depression, anxiety, and perceived stress are analyzed together (as dependent variables). Alternatively, one might analyze several subscales of depression, such as positive affect, negative affect, and somatic symptoms. MANOVA provides a convenience with a different type of omnibus test of all of the measures at once.

The null hypothesis tested with MANOVA is that all of the dependent variable means are equal. Because the algebraic equations become increasingly complex with multiple dependent variables, multivariate analysis are usually described in terms of matrices that summarize the multiple dependent measures. So, the null hypothesis is also a test of whether the vectors (columns) of means are equal across groups. A significant result indicates that one or more of the dependent-variable means differ among groups. Although usually a set of univariate ANOVA comparisons will be consistent with the MANOVA, there are some circumstances in which the results may be at odds with one another, typically because the MANOVA differences were somewhat weak and did not quite reach the significance level and because separate ANOVAs were more sensitive, focused tests of differences. A typical example would be when only one of several dependent measures shows a pattern of differences among the groups, resulting in an overall nonsignificant multivariate test but a significant univariate test for one specific measure.

Hotelling's T^2

There are several test statistics that are used with MANOVA (we had a brief introduction to them with repeated measures ANOVA). In the first of the multivariate test statistics, Hotelling (1931) developed a generalization of Gosset's t-test for the univariate case, now referred to as Hotelling's T^2. One way to state the univariate t-test is the following:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The difference between two group means, $\bar{y}_1$ and $\bar{y}_2$, in the numerator is evaluated relative to the standard error given in the denominator. In words, the standard error estimate is the square root of the weighted pooled variance divided by the group sample sizes and represents an estimate of sampling variability of the difference between the means. Pituch and Stevens (2016) show that, with a little algebra, the square of the t-test can be restated as

$$t^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{y}_1 - \bar{y}_2\right)\left(s^2\right)^{-1}\left(\bar{y}_1 - \bar{y}_2\right)$$

[1] Kesselman et al. (1998) make the argument that reducing Type I error is not good reason for conducting multivariate analysis of variance “There is very limited empirical support for this strategy. A counter position may be stated simply as follows: Do not conduct a MANOVA unless it is the multivariate effects that are of substantive interest. If the univariate effects are those of interest, then it is suggested that the researcher go directly to the univariate analyses and bypass MANOVA.” (pp. 361-362). Note also that the rationale to reduce familywise error and the need to use measures that are related and expected to show similar pattern of results are somewhat at odds with one another. To the extent that tests are redundant (e.g., measures are related) familywise error will be lower—it is maximal when tests are orthogonal.

Where s^2 is the pooled variance that combines s_1^2 and s_2^2 from above, and the superscript $^{-1}$ is the inverse (which is just $1/s^2$ here). Ignoring the first quantity on the right (the ratio of ns) for a minute, the three other quantities on the right represent a square of the mean differences $\left(\bar{y}_1 - \bar{y}_2\right)$ divided by the pooled variance. This restatement of the t-test is statistically equivalent to the Gosset equation for the first t equation given above, but it is convenient for expanding and expressing in terms of matrices that can contain more than one dependent variable.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)' \mathbf{S}^{-1}\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)$$

The bolded values of $\bar{\mathbf{y}}_1$ and $\bar{\mathbf{y}}_2$ represent column vectors of means (the $'$ symbol indicates the first parenthetical term is restated as a row), which contain the multiple dependent variables, and the $\mathbf{S}$ matrix is the variance-covariance matrix, containing variances for all of the y values on the diagonal and covariances on the off-diagonal. Hotelling's T^2 can be transformed to be evaluated against an F-distribution for significance and conceptualized in terms of a ratio of between-group to within-group mean squares.

Multivariate Statistical Tests

The Hotelling's T^2 is stated above in terms of two groups, but MANOVA can be used with any number of groups (i.e., levels of the independent variable), k, and any number of dependent variables, p. There are several highly related tests that are typically produced by MANOVA software procedures. Each of them can be stated in terms of the between-group and within-group variance. Below each is shown using $\mathbf{W}$ and $\mathbf{B}$, which are matrices of the sums of squares of y and their cross-products (the variance and covariance matrices without dividing by df). All of these measures are tested against an F-distribution for significance. When the result is significant, it indicates that differences existing among the groups on the dependent variables taken together. Wilks' lambda is the simplest and most straightforward in terms of its analogous relation to (the opposite of the) F ratio or eta-squared in univariate ANOVA, involving the ratio of mean square within to mean square total.

Wilks' lambda

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}$$

Hotelling's Trace

Hotelling's Trace (or Lawley-Hotelling trace) is a generalization of Hotelling's T^2, applying to k groups. The term "trace" comes from the matrix function that sums the diagonals of the matrix.

$$\frac{T^2}{N-k} \text{ or } \operatorname{trace}\left(\mathbf{E}^{-1}\mathbf{H}\right)$$

with *N* as total sample size, k as number of groups, $\mathbf{H}$ is the matrix of sum of squares cross-products for the hypothesis (explained) and $\mathbf{E}$ is the matrix sum of squares cross-products of errors.

Pillai's Trace

$$\operatorname{trace}\left[\mathbf{H}\left(\mathbf{H} + \mathbf{E}\right)^{-1}\right]$$

R*oy's Largest Root*

Roy's largest root (or sometimes Roy's greatest root) is the largest eigenvalue (see the "Principal Components" handout, covered next) from

$$\mathbf{E}^{-1}\mathbf{H}$$

With only two groups, Pillai's Trace, Wilks' Lambda, Hotellings Trace, and Roy's largest root are all equal. Pillai's Trace, Wilks' Lambda, Hotelling's Trace are asymptotically equivalent and will converge with larger samples. Olson (1976) suggests that Roy's largest root is too likely to produce Type I errors and should be avoided, that Wilks' lambda and Hotelling's trace are sensitive to violations of equal covariances in smaller samples, and he recommends Pillai's trace for general use.

Variance Accounted For
Owing to the close connection between Wilks' lambda and F from univariate ANOVA, we can compute eta-squared for the variance accounted for in the multivariate dependent variable composite by groups (Tabachnick & Fidell, 2013), is simply $\eta^2 = 1 - \Lambda$. Partial eta-squared is then $\eta^2 = 1 - \Lambda^{1/s}$, where s is based on the number of dependent variables, p, and degrees of freedom

$$s = \sqrt{\frac{p^2\left(df_{effect}\right)^2 - 4}{p^2 + \left(df_{effect}\right)^2 - 5}}$$

Assumptions
The assumptions for MANOVA overlap with those of standard ANOVA (and regression), such as independence of observations. There are somewhat stricter assumptions, however, and violations can lead to fairly severe increases in Type I or Type II errors (Christensen & Rencher, 1997; Coombs, Algina, & Oltman, 1996). Instead of the assumption that the single dependent variable is normally distributed in the population, MANOVA assumes the dependent variables together are multivariately normally distributed. This assumption is stricter in that, even if all the individual dependent variables are normally distributed, they may not be multivariate normally distributed. The analogy to the equal variance assumption is that the covariance matrices are equal across groups. Bartlett and Box tests can be used to make these comparisons, but they may be highly sensitive to minor departures with large N and insensitive to larger departures with small N (Huberty & Petoskey, 2000). There are a variety of proposed solutions to violations of these assumptions (see Coombs & Algina, 1996).

Other Possible Approaches
It is possible to stick to separate univariate ANOVAs and use a familywise error correction (such as the Sidak-Bonferroni correction; see the "Post Hoc Tests" handout for my univariate course). When the dependent variables are uncorrelated or weakly correlated, there may be little logical benefit to using MANOVA. MANOVA, in general, is a less powerful and a less focused test than univariate ANOVA (Tabachnick & Fidell, 2013).

Although MANOVA does not quite make an assumption that there is a single underlying construct explicitly, authors have cautioned against using MANOVA when the dependent variables are not related and are not expected to show similar results (Huberty & Petoskey, 2000). Otherwise, when the pattern of group differences is not the same across dependent variables, multivariate group differences will be weakened.[2] The MANOVA presumption that the dependent variables are assessing related or the same construct is never tested explicitly and could be in error. One simple alternative to MANOVA is to create a composite of the dependent variable and test group differences with a single (univariate) ANOVA. Composites should generally only be created when the variables are at least moderately correlated and are hypothesized to assess the same underlying construct. Such an analysis would differ from the MANOVA in two ways: (a) mean or summed composite indexes are usually equally weighted (e.g., measure 1 is not given more weight in the analysis than measure 2), whereas the MANOVA unequally weights the measures; and (b) in the single ANOVA of the composite, the composite is formed for the entire sample, whereas the MANOVA weights the measures maximizing the group differences. When the relative importance or weights of the items do not differ strongly and the correlation among the measures is high, one should not expect to find substantial differences between the composite ANOVA approach and the MANOVA approach.

[2] Negatively correlated dependent variables will work (Tabachnick & Fidell, 2013), so the dependent variables do not necessarily have to be recoded to be positively correlated for statistical reasons. Interpretation will be much simpler if the dependent variables are in the same direction, however.

A second, related approach would be to form an unequally weighted composite for the dependent variables based on an initial principal components analysis or factor analysis, approaches used to examine whether the underlying construct is unidimensional or not. Results from these initial analyses could then be used to form an unequally weighted composite and then analyzed with a univariate ANOVA (see Jackson, 1991, for a detailed discussion). This factor-scores approach usually results in a measure that is highly correlated with the unweighted composite (Fava & Velicer, 1992), and the univariate ANOVA results will nearly always lead to the same result whether factor scores are used or not. An advantage of the initial principal components or factor analysis is that the presumption about the unidimensionality of the set of dependent variables is explicitly tested.

A third approach would be to use structural equation modeling (SEM), where latent variables are formed from three or more measures assessing the same construct (a confirmatory factor model) and then groups are compared. A regression-based approach (e.g., with one or more grouping variables predicting the latent variable) or group mean comparisons of the latent variables can be tested (Thompson & Green, 2013). Confirmatory factor analysis may have some advantages over the principal components or exploratory factor analysis method in its sensitivity for selecting quality items and arriving at the most accurate factor structure (Brown, 2014), and the SEM approach has advantages in its capability to compare measurement properties across groups (testing for measurement "invariance"), incorporate more complex error structures, and expand to more complex models. The SEM approach allows the researcher to explore the equality of variances and covariances between groups for all variables individually or together but also does not require that either be equal across groups.

Extensions

The assigned reading (Pituch & Stevens, 2016) and the discussion above focus on a simple case in which two groups are compared, but MANOVA can be generalized to include more groups or multiple factors and to include covariates (MANCOVA). It is even possible to test multiple repeated measures with MANOVA (doubly multivariate), random effects, or block designs. MANOVA is also a simple case of regression analysis (see Timm, 2002, for proofs), so multivariate regression is also possible.

Follow-up post hoc tests can also be conducted (see Pituch & Stevens, 2016 for a more complete discussion). One test is the Roy-Bose simultaneous confidence interval approach. This is a generalization of the Scheffé test and is therefore equivalent to the overly conservative adjustments made by the Bonferroni correction. The other approaches used tend to employ the Bonferroni correction or a step-down or step-up procedure. Several of the alternative corrections, such as the Sidak-Bonferroni (discussed in the "Post Hoc Tests" handout in my univariate course) or step approaches (e.g., Hochberg's) can be adapted for specific-group for multiple unplanned multivariate tests or if follow-ups involve univariate comparisons planned contrasts or Tukey pairwise comparisons can be used.

References

Brown, T. A. (2014). *Confirmatory factor analysis for applied research, second edition.* New York: Guilford Publications.

Christensen, W. F., & Rencher, A. C. (1997). A comparison of Type I error rates and power levels for seven solutions to the multivariate Behrens-Fisher problem. *Communications in Statistics-Simulation and Computation, 26*, 1251-1273.

Coombs, W. T., & Algina, J. (1996). New test statistics for MANOVA/descriptive discriminant analysis. *Educational and Psychological Measurement, 56*(3), 382-402.

Coombs, W. T., Algina, J., & Oltman, D. O. (1996). Univariate and multivariate omnibus hypothesis tests selected to control Type I error rates when population variances are not necessarily equal. *Review of Educational Research, 66*, 137-179.

Fava, J. L., & Velicer, W. F. (1992). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research, 27*, 301-322.

Jackson, J. E. (2005). *A user's guide to principal components*. New York: John Wiley & Sons.

Huang, F. L. (2020). MANOVA: A procedure whose time has passed?. *Gifted Child Quarterly, 64*(1), 56-60.

Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In Tinsley, Howard EA, and Steven D. Brown, (eds.). *Handbook of applied multivariate statistics and mathematical modeling* (pp. 183-208). Academic Press.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*(3), 350-386.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin, 83*, 579-586.

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS.* New York: Routledge.

Tabachnick, B.G. and Fidell, L.S. (2013). *Using multivariate statistics, sixth edition.* Boston: Pearson.

Thompson, M.S., & Green, S.B. (2013). Evaluating between-group differences in latent variable means. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course, 2nd edition* (pp. 163-218). Charlotte, NC: Information Age Publishing.