

## Multivariate Analysis of Variance

What Multivariate Analysis of Variance is

The general purpose of multivariate analysis of variance (MANOVA) is to determine whether multiple levels of independent variables on their own or in combination with one another have an effect on the dependent variables. MANOVA requires that the dependent variables meet parametric requirements.

When do you need MANOVA?

MANOVA is used under the same circumstances as ANOVA but when there are multiple dependent variables as well as independent variables within the model which the researcher wishes to test. MANOVA is also considered a valid alternative to the repeated measures ANOVA when sphericity is violated.

What kinds of data are necessary?

The dependent variables in MANOVA need to conform to the parametric assumptions. Generally, it is better not to place highly correlated dependent variables in the same model for two main reasons. First, it does not make scientific sense to place into a model two or three dependent variables which the researcher knows measure the same aspect of

outcome. (However, this is point will be influenced by the hypothesis which the researcher is testing. For example, subscales from the same questionnaire may all be included in a MANOVA to overcome problems associated with multiple testing. Subscales from most questionnaires are related but may represent different aspects of the dependent variable.) The second reason for trying to avoid including highly correlated dependent variables is that the correlation between them can reduce the power of the tests. If MANOVA is being used to reduce multiple testing, this loss in power needs to be considered as a trade-off for the reduction in the chance of a Type I error occurring.

Homogeneity of variance from ANOVA and t tests becomes homogeneity of variance covariance in MANOVA models. The amount of variance within each group needs to be comparable so that it can be assumed that the groups have been drawn from a similar population. Furthermore it is assumed that these results can be pooled to produce an error value which is representative of all the groups in the analysis. If there is a large difference in the amount of error within each group the estimated error measure for the model will be misleading.

How much data?

There needs to be more participants than dependent variables. If there were only one participant in any one of the combination of conditions, it would be impossible to determine the amount of variance within that combination (since only one data point would be available). Furthermore, the statistical power of any test is limited by a small sample size. (A greater amount of variance will be attributed to error in smaller sample

sizes, reducing the chances of a significant finding.) A value known as Box's M, given by most statistical programs, can be examined to determine whether the sample size is too small. Box's M determines whether the covariance in different groups is significantly different and must not be significant if one wishes to demonstrate that the sample sizes in each cell are adequate. An alternative is Levene's test of homogeneity of variance which tolerates violations of normality better than Box's M. However, rather than directly testing the size of the sample it examines whether the amount of variance is equally represented within the independent variable groups.

In complex MANOVA models the likelihood of achieving robust analysis is intrinsically linked to the sample size. There are restrictions associated with the generalizability of the results when the sample size is small and therefore researchers should be encouraged to obtain as large a sample as possible.

### Example of MANOVA

Considering an example may help to illustrate the difference between ANOVAs and MANOVAs. Kaufman and McLean (1998) used a questionnaire to investigate the relationship between interests and intelligence. They used the Kaufman Adolescent and Adult Intelligence Test (KAIT) and the Strong Interest Inventory (SII) which contained six subscales on occupational themes (GOT) and 23 Basic Interest Scales (BISs). Kaufman et al. used a MANOVA model which had four independent variables: age, gender, KAIT IQ and Fluid-Crystal intelligence (F-C). The dependent variables were the six occupational theme subscales (GOT) and the twenty-three Basic Interest Scales (BIS).

In Table 2.1 the dependent variables are listed in columns 3 and 4. The independent variables are listed in column 2, with the increasingly complex interactions being shown below the main variables.

If an ANOVA had been used to examine these data, each of the GOT and BIS subscales would have been placed in a separate ANOVA. However, since the GOT and BIS scales are related, the results from separate ANOVAs would not be independent. Using multiple ANOVAs would increase the risk of a Type I error (a significant finding which occurs by chance due to repeating the same test a number of times).

Kaufman and McLean used the Wilks' lambda multivariate statistic (similar to the F values in univariate analysis) to consider the significance of their results and reported only the interactions which were significant. These are shown as Sig in Table 2.1. The values which proved to be significant are the majority of the main effects and one of the 2-way interactions. Note that although KAIT IQ had a significant main effect none of the interactions which included this variable were significant. On the other hand, age and gender show a significant interaction in the effect which they have on the dependent variables.

What a multivariate analysis of variance does

Like an ANOVA, MANOVA examines the degree of variance within the independent variables and determines whether it is smaller than the degree of variance between the independent variables. If the within subjects variance is smaller than the between subjects variance it means the independent variable has had a significant effect on the dependent

Table 2.1. The different aspects of the data considered by the MANOVA model used by Kaufman and McLean (1998).

Level	Independent variables	6 GOT subscales	23 BIS
Main Effects	Age		Sig
	Gender	Sig	Sig
	KAIT IQ	Sig	Sig
	F-C		
2-way Interactions	Age x Gender	Sig	Sig
	Age x KAIT IQ		
	Age x F-C		
	Gender x KAIT IQ		
	Gender x F-C		
	KAIT IQ x F-C		
3-way Interactions	Age x Gender x KAIT IQ		
	Age x Gender x F-C		
	Age x KAIT IQ x F-C		
	Gender x KAIT IQ x F-C		
	Age KAIT IQ x F-C		
4-way Interactions	Age x Gender x KAIT IQ x F-C		

variables. There are two main differences between MANOVAs and ANOVAs. The first is that MANOVAs are able to take into account multiple independent and multiple dependent variables within the same model, permitting greater complexity. Secondly, rather than using the F value as the indicator of significance a number of multivariate measures (Wilks' lambda, Pillai's trace, Hotelling trace and Roy's largest root) are used. (An explanation of these multivariate statistics is given below).

MANOVA deals with the multiple dependent variables by combining them in a linear manner to produce a combination which best separates the independent variable groups. An ANOVA is then performed on the newly developed dependent variable. In MANOVAs the independent variables relevant to each main effect are weighted to give them priority in the calculations performed. In interactions the independent variables are equally weighted to determine whether or not they have an additive effect in terms of the combined variance they account for in the dependent variable/s.

The main effects of the independent variables and of the interactions are examined with "all else held constant". The effect of each of the independent variables is tested separately. Any multiple interactions are tested separately from one another and from any significant main effects. Assuming there are equal sample sizes both in the main effects and the interactions, each test performed will be independent of the next or previous calculation (except for the error term which is calculated across the independent variables).

There are two aspects of MANOVAs which are left to researchers: first, they decide which variables are placed in the MANOVA. Variables are included in order to address a

particular research question or hypothesis, and the best combination of dependent variables is one in which they are not correlated with one another, as explained above. Second, the researcher has to interpret a significant result. A statistical main effect of an independent variable implies that the independent variable groups are significantly different in terms of their scores on the dependent variable. (But this does not establish that the independent variable has caused the changes in the dependent variable. In a study which was poorly designed, differences in dependent variable scores may be the result of extraneous, uncontrolled or confounding variables.)

To tease out higher level interactions in MANOVA, smaller ANOVA models which include only the independent variables which were significant can be used in separate analyses and followed by post hoc tests. Post hoc and preplanned comparisons compare all the possible paired combinations of the independent variable groups e.g. for three ethnic groups of white, African and Asian the comparisons would be: white v African, white v Asian, African v Asian. The most frequently used preplanned and post hoc tests are Least Squares Difference (LSD), Scheffe, Bonferroni, and Tukey. The tests will give the mean difference between each group and a p value to indicate whether the two groups differ significantly.

The post hoc and preplanned tests differ from one another in how they calculate the p value for the mean difference between groups. Some are more conservative than others. LSD perform a series of t tests only after the null hypothesis (that there is no overall difference between the three groups) has been rejected. It is the most liberal of the post hoc tests and has a high Type I error rate. The Scheffe test uses the F distribution rather than the t distribution of the LSD tests and is considered more conservative. It has a high

Type II error rate but is considered appropriate when there are a large number of groups to be compared. The Bonferroni approach uses a series of t tests but corrects the significance level for multiple testing by dividing the significance levels by the number of tests being performed (for the example given above this would be  $0.05/3$ ). Since this test corrects for the number of comparisons being performed, it is generally used when the number of groups to be compared is small. Tukey's Honest Significance Difference test also corrects for multiple comparisons, but it considers the power of the study to detect differences between groups rather than just the number of tests being carried out i.e. it takes into account sample size as well as the number of tests being performed. This makes it preferable when there are a large number of groups being compared, since it reduces the chances of a Type I error occurring.

The statistical packages which perform MANOVAs produce many figures in their output, only some of which are of interest to the researcher.

**Sum of Squares:** The sum of squares measure found in a MANOVA, like that reported in the ANOVA, is the measure of the squared deviations from the mean both within and between the independent variable. In MANOVA, the sums of squares are controlled for covariance between the independent variables.

There are six different methods of calculating the sum of squares. Type I, hierarchical or sequential sums of squares, is appropriate when the groups in the MANOVA are of equal sizes. Type I sum of squares provides a breakdown of the sums of squares for the whole model used in the MANOVA but it is particularly sensitive to the order in which the independent variables are placed in the model. If a variable is entered first, it is not

adjusted for any of the other variables; if it is entered second, it is adjusted for one other variable (the first one entered); if it is placed third, it will be adjusted for the two other variables already entered.

Type II, the partially sequential sum of squares, has the advantage over Type I in that it is not affected by the order in which the variables are entered. It displays the sum of squares after controlling for the effect of other main effects and interactions but is only robust where there are even numbers of participants in each group.

Type III sum of squares can be used in models where there are uneven group sizes, although there needs to be at least one participant in each cell. It calculates the sum of squares after the independent variables have all been adjusted for the inclusion of all other independent variables in the model.

Type IV sum of squares can be used when there are empty cells in the model but it is generally thought more suitable to use Type III sum of squares under these conditions since Type IV is not thought to be good at testing lower order effects.

Type V has been developed for use where there are cells with missing data. It has been designed to examine the effects according to the degrees of freedom which are available and if the degrees of freedom fall below a given level these effects are not taken into account. The cells which remain in the model have at least the degrees of freedom the full model would have without any cells being excluded. For those cells which remain in the model the Type III sum of squares are calculated. However, the Type V sum of squares

are sensitive to the order in which the independent variables are placed in the model and the order in which they are entered will determine which cells are excluded.

Type VI sum of squares is used for testing hypotheses where the independent variables are coded using negative and positive signs e.g. +1 = male, -1 = female.

Type III sum of squares is the most frequently used as it has the advantages of Types IV, V and VI without the corresponding restrictions.

Mean Squares: The mean square is the sum of squares divided by the appropriate degrees of freedom.

Multivariate Measures: In most of the statistical programs used to calculate MANOVAs there are four multivariate measures: Wilks' lambda, Pillai's trace, Hotelling-Lawley trace and Roy's largest root. The difference between the four measures is the way in which they combine the dependent variables in order to examine the amount of variance in the data. Wilks' lambda demonstrates the amount of variance accounted for in the dependent variable by the independent variable; the smaller the value, the larger the difference between the groups being analyzed.  $1 - \text{Wilks' lambda}$  indicates the amount of variance in the dependent variables accounted for by the independent variables. Pillai's trace is considered the most reliable of the multivariate measures and offers the greatest protection against Type I errors with small sample sizes. Pillai's trace is the sum of the variance which can be explained by the calculation of discriminant variables. It calculates the amount of variance in the dependent variable which is accounted for by the greatest separation of the independent variables. The Hotelling-Lawley trace is generally

converted to the Hotelling's T-square. Hotelling's T is used when the independent variable forms two groups and represents the most significant linear combination of the dependent variables. Roy's largest root, also known as Roy's largest eigenvalue, is calculated in a similar fashion to Pillai's trace except it only considers the largest eigenvalue (i.e. the largest loading onto a vector). As the sample sizes increase the values produced by Pillai's trace, Hotelling-Lawley trace and Roy's largest root become similar. As you may be able to tell from these very broad explanations, the Wilks' lambda is the easiest to understand and therefore the most frequently used measure.

**Multivariate F value:** This is similar to the univariate F value in that it is representative of the degree of difference in the dependent variable created by the independent variable. However, as well as being based on the sum of squares (as in ANOVA) the calculation for F used in MANOVAs also takes into account the covariance of the variables.

### Example of Output of SPSS

The data analysed in this example came from a large-scale study with three dependent variables which were thought to measure distinct aspects of behaviour. The first was total score on the Dissociative Experiences Scale, the second was reaction time when completing a signal detection task and the third was a measure of people's hallucinatory experiences (the Launay Slade Hallucinations Scale, LSHS). The independent variable was the degree to which participants were considered prone to psychotic experiences (labeled PRONE in the output shown in Fig 2.1), divided into High, Mean and Low.

The first part of the output is shown in Fig 2.1. The Between-Subjects Factors table displays the independent variable levels. Here there is only one independent variable with three levels. The number of participants in each of the independent variable groups are displayed in the column on the far right.

The Multivariate Tests table displays the multivariate values: Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root. These are the multivariate values for the model as a whole. The F values for the Intercept, shown in the first part of the table are all the same. Those for the independent variable labeled PRONE are all different, but they are all significant above the 1% level (shown by Sig being .000), indicating that on the dependent variables there is a significant difference between the three proneness groups.

The Tests of Between-Subjects Effects table gives the sum of squares, degrees of freedom, Mean Square value, the F values and the significance levels for each dependent variable. (The Corrected Model is the variance in the dependent variables which the independent variables accounts for without the intercept being taken into consideration.) The section of the table which is of interest is where the source under consideration is the independent variable, the row for PRONE. In this row it can be seen that two (DES, Launay Slade Hallucinations Scale) out of the three dependent variables included in the model are significant ( $p < .05$ ), meaning that three proneness groups differ significantly in their scores on the DES and the LSHS.

\*\*\*\*\* Insert Fig 2.1 \*\*\*\*\*

Figure 2.1. Example of main output from SPSS for MANOVA.

To determine which of the three independent variable groups differ from one another on the dependent variables, Least Squares difference comparisons were performed by selecting the relevant option in SPSS. The output is shown in Fig 2.2. The table is split into broad rows, one for each dependent variable, and within each row the three groups of high, mean or low Proneness are compared one against the other. The mean difference between the two groups under consideration are given in one column and then separate columns show the Standard Error, Significance and the lower and upper 95% Confidence intervals. For the DES dependent variable, the High group is significantly different from the Low group ( $p=0.004$ ) but not from the Mean group ( $p=0.076$ ). Similarly, for the Total mean reaction time dependent variable, the High group differs significantly from the Low group but not from the Mean group. For both these dependent variables, the Mean group do not differ from the High group nor from the Low group. For the Launay scale, the High group differs significantly from the Mean and Low groups, both of which also differ from each other.

\*\*\*\*\* Insert Fig 2.2 \*\*\*\*\*

Figure 2.2. The Least Squares Difference output from SPSS MANOVA analysis.

The means and standard deviations on the DES and Launay scales for the three proneness groups are displayed in Fig 2.3, which was produced by SPSS. Tables such as this assist in interpreting the results.

The MANOVA above could also have included another independent variable, such as gender. The interaction between the two independent variables on the dependent variables would have been reported in the Multivariate Statistics and Tests of Between Subjects Effects tables.

\*\*\*\*\* Insert Fig 2.3 \*\*\*\*\*

Figure 2.3. Means for the three psychosis proneness groups on the Dissociative Experiences Scale and the Launay Slade Hallucinations Scale.

Example of the use of MANOVA

From health

Snow and Bruce (2003) explored the factors involved in Australian teenage girls smoking, collecting data from 241 participants aged between 13 and 16 years of age. Respondents completed a number of questionnaires including an Adolescent Coping Scale. They were also asked to indicate how frequently they smoked cigarettes and the responses were used to divide the respondents into three smoking groups (current smokers, experimental smokers, never smokers). In their analysis, Snow and Bruce used

MANOVA with smoking group as the independent variable. In one of the MANOVAs, the dependent variables were the measures on three different coping strategies. Snow and Bruce were only interested in the main effects of smoking group on the dependent variables so they converted the Wilk's lambda to F values and significance levels. They used Scheffe post hoc analysis to determine which of the three smoking groups differed significantly on the dependent variables, and found that on the 'productive' coping strategy there was a significant difference between the current and experimental smokers, on the 'non-productive' coping strategy there was a difference between the current smokers and those who had never smoked, and on the 'rely on others' coping strategy there was a difference between current and experimental smokers.

## FAQs

How do I maximise the power in a MANOVA?

Power refers to the sensitivity of your study design to detect true significant findings when using statistical analysis. Power is determined by the significance level chosen, the effect size and the sample size. (There are many free programmes available on the internet which can be used to calculate effect sizes from previous studies by placing means and sample sizes into the equations.) A small effect size will need a large sample size for significant differences to be detected, while a large effect size will need a relatively small sample to be detected.

In general, the power of your analysis will increase the larger the effect size and sample size. Taking a practical approach, obtaining as large a sample as possible will maximise the power of your study.

What do I do if I have groups with uneven sample sizes?

Having unequal sample sizes can affect the integrity of the analysis. One possibility is to recruit more participants into the groups which are under represented in the data set, another is to randomly delete cases from the more numerous groups until they are equal to the least numerous one.

What do I do if I think there may be a covariate with the dependent variables included in a MANOVA model?

When a covariate is incorporated into a MANOVA it is usually referred to as a MANCOVA model. The 'best' covariate for inclusion in a model should be highly correlated with the dependent variables but not related to the independent variables. The dependent variables included in a MANCOVA are adjusted for their association with the covariate. Some experimenters include baseline data in as a covariate to control for any individual differences in scores since even randomisation to different experimental conditions does not completely control for individual differences.

Summary

MANOVA is used when there are multiple dependent variables as well as independent variables in the study. MANOVA combines the multiple dependent variables in a linear manner to produce a combination which best separates the independent variable groups. An ANOVA is then performed on the newly developed dependent variable.

## Glossary

**Additive:** the effect of the independent variables on one another when placed in multi- or univariate analysis of variance.

**Interaction:** the combined effect of two or more independent variables on the dependent variables.

**Main Effect:** the effect of one independent variable on the dependent variables, examined in isolation from all other independent variables.

**Mean Squares:** sum of squares expressed as a ratio of the degrees of freedom either within or between the different groups.

**Sum of Squares:** the squared deviations from the mean within or between the groups. In MANOVA they are controlled for covariance.

**Wilks' lambda:** a statistic which can vary between 0 and 1 and which indicates whether the means of groups differ. A value of 1 indicates the groups have the same mean.

## References

Kaufman, A.S. and McLean, J.E. (1998). An investigation into the relationship between interests and intelligence. *Journal of Clinical Psychology*, 54, 279-295.

Snow, P.C. and Bruce, D.D. (2003). Cigarette smoking in teenage girls: exploring the role of peer reputations, self-concept and coping. *Health Education Research Theory and Practice*, 18, 439-452.

### Further Reading

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Rees, D.G. (2000). *Essential Statistics*. (4th ed.) London: Chapman and Hall/CRC.

### Internet Sources

[www.statsoftinc.com/textbook/](http://www.statsoftinc.com/textbook/)

[www2.chass.ncsu.edu/garson/pa765/manova.htm](http://www2.chass.ncsu.edu/garson/pa765/manova.htm)