

# MANOVA and MANCOVA

## 1 Introduction

MANOVA (Multivariate Analysis of Variance) and MANCOVA (Multivariate Analysis of Covariance) are used when we want to analyse differences in multiple dependent variables across groups.

- MANOVA extends ANOVA by examining the effect of one or more independent variables on multiple correlated dependent variables, considering their interrelationships.
- MANCOVA further incorporates covariates to control for confounding variables, increasing the precision of group comparisons by accounting for extraneous variability.

They help us explore complex relationships while maintaining statistical power.

### 1.1 Example

Imagine an analyst wants to study the effect of different training programs (independent variable) on athletes' performance. They measure two dependent variables: speed and endurance.

- MANOVA helps determine if the training programs have a statistically significant effect on these correlated performance metrics.

Building on this example, the coach includes the athletes' age as a covariate to control for its influence on performance.

- MANCOVA assesses whether differences in speed and endurance are due to the training programs, independent of age effects.

## 2 Multivariate Analysis of Variance (MANOVA)

### 2.1 Introduction

Multivariate Analysis of Variance (MANOVA) extends the capabilities of ANOVA by allowing us to analyse multiple dependent variables simultaneously.

It helps determine how different groups vary across several measures, taking into account the relationships between dependent variables.

Unlike univariate ANOVA, which examines one dependent variable at a time, MANOVA considers the interplay between multiple dependent variables. This gives us a more comprehensive understanding of group differences while controlling for the relationship between outcome variables.

Key advantages of MANOVA include:

- Reduced Type I error rate compared to running multiple separate ANOVAs;
- Ability to detect patterns that might not be evident when analysing variables separately; and
- More powerful analysis when dependent variables are moderately correlated.

### 2.2 Dependent variables and MANOVA

In Multivariate Analysis of Variance, selecting appropriate dependent variables is really important. Instead of analysing a single outcome (like ANOVA), MANOVA considers multiple interrelated measures as a single multivariate response, using their correlations to better assess group differences.

A key advantage of MANOVA is reducing Type I error inflation. If we run separate ANOVAs on correlated outcomes increases the risk of false significance. By evaluating the joint distribution of dependent variables, MANOVA consolidates tests into one framework, offering us a clearer picture of how outcomes interact, particularly when strong correlations exist.

Choosing dependent variables requires careful justification. Each variable should add unique insight while maintaining enough overlap for joint analysis. Redundant variables contribute little, while overly diverse ones weaken our model's focus. A well-chosen set enhances understanding of the phenomenon within a coherent statistical structure.

## 2.3 Independent variables and MANOVA

Independent variables (or factors) in a MANOVA framework typically represent categorical treatments, group memberships, or experimental conditions.

Their role is to partition the multidimensional space of the dependent variables, allowing us to test whether different levels of these factors produce distinct profiles of outcomes. Much like univariate ANOVA, these factors can be crossed or nested, enabling complex designs that map neatly onto multidimensional response surfaces.

A critical consideration in MANOVA is the assumption of multivariate normality and homogeneity of the variance-covariance matrices across groups. Violations of these assumptions can compromise the validity of the statistical tests and lead to incorrect inferences. Therefore, we would usually perform diagnostic checks (e.g., Box's M test) and consider robust or nonparametric alternatives if our data substantially deviate from these requirements.

Whether dealing with between-subjects, within-subjects (repeated measures), or mixed designs, it's important to structure the analysis in a way that captures the complexity of the experimental setup. MANOVA offers the flexibility to investigate main effects, interactions, and higher-order terms within a single model, all while retaining information about how the dependent variables relate to one another.

Ultimately, the effect of these independent variables can be interpreted through both omnibus tests of overall multivariate differences and follow-up evaluations of specific contrasts or linear combinations of dependent measures.

## 3 Assumptions of MANOVA

### 3.1 Introduction

Remember: MANOVA is used when we want to analyse group differences across multiple dependent variables simultaneously.

As we've noted, unlike ANOVA (which examines one dependent variable at a time) MANOVA considers the relationships between dependent variables and helps control for Type I error that might occur with multiple separate ANOVAs.

Just like all statistical tests, before conducting a MANOVA several key assumptions must be met to ensure the validity of the results. However, these assumptions are a bit more complex than those for univariate analyses due to the multivariate nature of the test (which is also its strength).

## 3.2 Assumption 1: Multivariate normality

### ! Important

The assumption of multivariate normality is fundamental to MANOVA. This requires that all dependent variables collectively follow a multivariate normal distribution.

This means that not only should each dependent variable be normally distributed individually, but their combinations should also follow a normal distribution pattern!

This assumption becomes particularly important when working with multiple dependent variables simultaneously, as it affects the reliability of test statistics and p-values in MANOVA.

### 3.2.1 Testing for multivariate normality

Testing for multivariate normality can be more complex than testing for univariate normality. While individual variables can be assessed using traditional methods like Q-Q plots and Shapiro-Wilk tests, multivariate normality often requires more sophisticated approaches such as Mardia's test or the Henze-Zirkler test.

Fortunately, MANOVA is relatively robust to moderate violations of this assumption, especially when sample sizes are large and groups are of approximately equal size. [1], [2]

## 3.3 Assumption 2: Homogeneity of variance-covariance matrices

### ! Important

Homogeneity of variance-covariance matrices is another critical assumption in MANOVA.

This assumption requires that the relationships between dependent variables and their variability remain consistent across all groups being compared.

When this assumption is met, we can be confident that any differences found between groups are not simply due to differences in how the variables relate to each other within each group.

Violation of this assumption can lead to distorted results and incorrect conclusions, especially when group sizes are unequal. If the assumption is violated with larger sample sizes, Pillai's trace criterion (see below) is usually recommended, as it is more robust.

However, when group sizes are equal, the analysis is generally robust to violations of this assumption, though interpretation should still proceed with caution. [2]

### 3.3.1 Testing for homogeneity

The homogeneity of variance-covariance matrices can be tested using Box's M test, which examines whether the covariance matrices are equal across groups.

Steps to conduct Box's M test:

1. Calculate the covariance matrices for each group separately
2. Compare these matrices to test if they are statistically similar
3. Examine the p-value: if  $p > .001$ , the assumption is typically considered met

Note: Box's M test is highly sensitive to violations of normality and to large sample sizes. Because of this sensitivity, many researchers use a more conservative alpha level ( $p < .001$ ) rather than the conventional .05.

Alternative approaches include:

- Visual inspection of variance-covariance matrices
- Examining the ratio of largest to smallest variance (should be less than 10:1)
- Using Levene's test for individual variables

## 3.4 Assumption 3: Linear relationships

### ! Important

MANOVA also assumes that dependent variables are linearly related to each other and to the independent variables.

This assumption is critical because MANOVA relies on linear combinations of the dependent variables to assess group differences effectively. If the relationships are non-linear, the model may fail to capture meaningful patterns, reducing statistical power and leading to misleading results.

A strong linear relationship among dependent variables ensures that MANOVA can detect differences between groups based on meaningful multivariate patterns. If the relationships are weak or non-linear, the technique may not provide a reliable summary of the data, as the underlying structure is not well represented by the model. Checking for linearity is therefore essential before proceeding with the analysis.

To assess this assumption, researchers can use scatterplots, correlation matrices, or statistical tests such as Bartlett's test of sphericity. If significant non-linearity is detected, possible solutions include transforming the data or considering alternative methods, such as non-parametric techniques or Generalized Additive Models (GAMs), which allow for non-linear relationships.

## 4 Multivariate Analysis of Covariance (MANCOVA)

### 4.1 Introduction

Multivariate Analysis of Covariance (MANCOVA) extends MANOVA by incorporating one or more *covariates* into the analysis.

A “covariate” is a continuous variable that is included in a statistical model to account for its potential influence on the dependent variable, helping to control for confounding effects and improve the accuracy of the analysis. Age is often included as a covariate in sport analytics.

It allows us to examine the effects of categorical independent variables on multiple dependent variables while controlling for continuous variables (‘covariates’) that might influence the relationship.

MANCOVA extends MANOVA by including one or more covariates. These are variables that are not of primary interest but may influence the outcome.

By controlling for these covariates, MANCOVA can reduce the error variance and increase the sensitivity of the test for detecting true differences.

#### 4.1.1 What’s it used for?

MANCOVA is particularly useful when:

- We have multiple dependent variables that are correlated;
- We need to **control** for variables that could affect our results; or
- We want to increase statistical power by reducing error variance.

Like its univariate counterpart ANCOVA, MANCOVA helps reduce bias in our results by accounting for pre-existing differences between groups.

However, it does this across multiple dependent variables simultaneously, making it a more comprehensive analytical tool for complex research designs.

### 4.2 Covariates

In sports research, covariates are continuous variables that help control for factors that might influence performance outcomes, but are not of primary interest.

For example, in a study examining the impact of different training programmes on sprint speed, variables such as age, baseline fitness level, or muscle mass could act as covariates to

ensure that observed differences in speed are due to training rather than pre-existing physical advantages.

By including covariates in statistical models like MANCOVA, we can adjust for these influences, leading to more accurate comparisons between groups and a clearer understanding of the true effect of the independent variable on performance.

### 4.3 Adjusted means

MANCOVA calculates group means adjusted for the effects of covariates.

“Adjusted means” refer to group means that have been statistically modified to account for the influence of covariates, providing a clearer comparison of group effects.

In an analysis where a covariate, such as age, is controlled for, the adjusted means represent what the group means would be if all participants were equal on that covariate.

This adjustment helps eliminate bias that could arise if one group had systematically higher or lower values of the covariate, ensuring that differences between groups are not confounded by external factors.

Adjusted means are particularly useful in studies where groups cannot be perfectly balanced across key variables

### 4.4 Summary

Multivariate Analysis of Covariance (MANCOVA) extends MANOVA by incorporating covariates to control for their influence on multiple dependent variables simultaneously.

It's particularly useful when we suspect that an extraneous variable may affect all dependent measures and want to statistically remove its impact to better isolate the effect of the independent variable.

For example, in a study comparing the psychological and physiological effects of different coaching styles, a covariate like prior experience could be included to ensure that differences in motivation and endurance are attributed to coaching rather than past training history.

MANCOVA enhances the precision of group comparisons by reducing error variance and improving the validity of conclusions drawn from multivariate data.

## 5 Interpreting MANOVA and MANCOVA Models

MANOVA and MANCOVA are interpreted through a systematic approach that

1. examines multivariate test statistics to determine overall group differences, then;
2. conducts univariate analyses to identify specific differences between groups on individual dependent variables.

The interpretation considers both statistical significance and effect sizes to understand the practical importance of findings.

Key aspects of interpretation include:

- Evaluating overall multivariate effects using test statistics;
- Examining individual variable contributions through univariate tests;
- Assessing the magnitude of effects through effect size measures; and
- Considering covariate adjustments (in MANCOVA).

### 5.1 Multivariate tests in MANOVA and MANCOVA

Multivariate tests in MANOVA and MANCOVA evaluate whether group means differ across multiple dependent variables. The most common test statistics—**Pillai’s Trace**, **Wilk’s Lambda**, and **Hotelling’s Trace**—offer slightly different perspectives on the same hypothesis. *Pillai’s Trace* is the most robust to violations of assumptions, making it a preferred choice when there are concerns about multivariate normality or homogeneity of covariance. *Wilk’s Lambda*, the most widely used, measures the proportion of total variance not explained by group differences, with smaller values indicating stronger group effects.

*Hotelling’s Trace*, conceptually similar to Pillai’s Trace, is more powerful when groups are well-separated but less robust to assumption violations.

When interpreting these statistics, researchers assess their associated *p-values* to determine statistical significance. If a test is significant, it suggests that at least one dependent variable differs across groups. However, significant results do not indicate which specific variables contribute to the effect—this requires follow-up univariate tests.

To illustrate how these statistics are computed in R, we can perform MANOVA on the `iris` dataset, treating species as the independent variable and the four flower measurements as dependent variables.

```
              Df Pillai approx F num Df den Df      Pr(>F)
Species      2 1.1919   53.466      8   290 < 2.2e-16 ***
Residuals 147
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

          Df      Wilks approx F num Df den Df      Pr(>F)
Species    2 0.023439   199.15      8    288 < 2.2e-16 ***
Residuals 147

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

          Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
Species    2      32.477   580.53      8    286 < 2.2e-16 ***
Residuals 147

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 5.2 Univariate tests

After a significant MANOVA result, univariate tests (essentially ANOVAs for each dependent variable) help pinpoint which variables contribute to group differences. While MANOVA controls for Type I error inflation across multiple dependent variables, individual ANOVAs provide more specific insights into how each variable differs. In MANCOVA, these tests are adjusted for covariates, ensuring that observed differences are not confounded by extraneous variables.

The interpretation of univariate tests follows standard ANOVA principles: each dependent variable is tested separately for group differences using F-tests. If multiple univariate tests are conducted, applying Bonferroni or Holm corrections helps control for inflated error rates. Below, we extract and visualize univariate test results from the MANOVA model:

```

Response Sepal.Length :
          Df Sum Sq Mean Sq F value      Pr(>F)
Species    2 63.212  31.606  119.26 < 2.2e-16 ***
Residuals 147 38.956   0.265

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Response Sepal.Width :
          Df Sum Sq Mean Sq F value      Pr(>F)
Species    2 11.345   5.6725   49.16 < 2.2e-16 ***
Residuals 147 16.962   0.1154

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Response Petal.Length :
          Df Sum Sq Mean Sq F value      Pr(>F)

```

```
Species      2 437.10 218.551 1180.2 < 2.2e-16 ***
Residuals   147 27.22 0.185
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

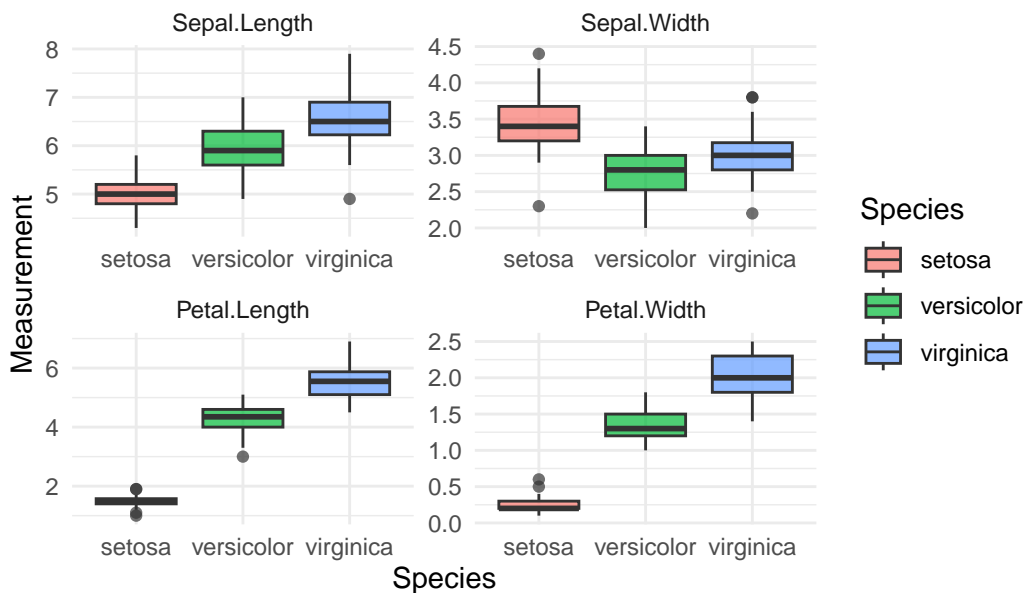
Response Petal.Width :

```
      Df Sum Sq Mean Sq F value    Pr(>F)
Species  2 80.413  40.207  960.01 < 2.2e-16 ***
Residuals 147  6.157   0.042
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Univariate Boxplots for Dependent Variables



### 5.3 Effect sizes

As we noted earlier in the module, statistical significance alone does not indicate the practical importance of findings. Effect sizes quantify the strength of group differences, providing a clearer understanding of their real-world impact.

In MANOVA and MANCOVA, the most common effect size measure is partial eta-squared ( $\eta^2$ ), which represents the proportion of variance in each dependent variable explained by the independent variable. Wilk's Lambda, Pillai's Trace, and Hotelling's Trace can also be converted into an effect size metric called generalized eta-squared ( $\eta^2_g$ ) for multivariate interpretation.

Effect sizes are particularly valuable in MANCOVA, where adjusted means may reduce the observed variance, making statistical significance harder to achieve. Even if p-values are marginal, strong effect sizes indicate meaningful relationships.

A larger partial eta-squared (typically above 0.14) suggests a strong effect, while values below 0.06 indicate a weak effect. Understanding effect sizes ensures that findings are not only statistically significant but also substantively meaningful.

## 6 Applications and Limitations

### 6.1 Introduction

As we've noted, MANOVA and MANCOVA are powerful statistical techniques with specific applications and limitations:

- MANOVA allows analysis of multiple dependent variables in a single test
- MANCOVA extends this by including covariates to control for external influences
- Both methods help reduce *Type I* error compared to multiple individual tests
- They require careful consideration of sample size and assumptions

### 6.2 Some applications in sport

Here are some examples where the application of MANOVA and MANCOVA can be highly useful within a sport data analytics context:

#### Performance Analysis in Sport

MANOVA is frequently applied in sports science to assess the impact of training interventions across multiple performance metrics. For example, a study might evaluate how a high-intensity training program affects sprint speed, agility, and vertical jump height in athletes. By analysing these dependent variables simultaneously, MANOVA can reveal whether the training regimen has a global or domain-specific effect on performance.

#### Injury Prevention and Rehabilitation

MANCOVA is particularly valuable in accounting for pre-existing conditions or baselines when evaluating interventions. For instance, in a study of rehabilitation programs for ACL injuries, researchers could analyse improvements in strength, range of motion, and functional performance, while using pre-injury fitness levels as covariates to control for individual differences.

#### Talent Identification and Development

MANCOVA is useful in talent development programs where baseline characteristics such as age or training history might confound outcomes. For example, in analyzing the success

of a youth development initiative, MANCOVA could control for training experience while evaluating growth in physical metrics (e.g., VO2 max, body composition) and skill acquisition (e.g., passing accuracy, decision-making speed).

### 6.3 Limitations of MANOVA and MANCOVA

MANOVA and MANCOVA rely on strict statistical assumptions, including *multivariate normality*, *homogeneity of covariance matrices* (tested by Box's M), and *linear relationships* between covariates and dependent variables.

Violating these assumptions can produce biased or invalid results. For example, skewed data or extreme outliers (common in elite athlete performance), can reduce the robustness of these techniques.

Both methods are complex and require a degree of expertise, particularly when interpreting interactions or high-dimensional datasets. In sports science, common dependent variables like physical performance metrics may interact unpredictably or be influenced by confounding factors, complicating our analysis.

Large sample sizes are often necessary, especially when multiple dependent variables are involved. This poses challenges in sports studies with elite athletes, where small sample sizes can lead to underpowered analyses and reduced generalisability.

While MANOVA and MANCOVA offer a useful approach to analysis, their findings do not always translate well to practical applications.

Significant multivariate effects observed in controlled settings may not account for dynamic real-world influences during competition. Additionally, these methods can overemphasise statistical significance. Small effects may be deemed significant, but their practical impact on performance or outcomes could be minimal.

### 6.4 Extensions of MANOVA and MANCOVA

Techniques have been developed that represent powerful extensions of MANOVA (Multivariate Analysis of Variance) and MANCOVA (Multivariate Analysis of Covariance).

These techniques can offer deeper insights and more nuanced analysis capabilities, making them particularly valuable in complex research scenarios.

### 6.4.1 Discriminant Function Analysis (DFA)

Following MANOVA, DFA can identify which dependent variables contribute most to the observed group differences. For instance, in talent identification, DFA might highlight that sprint speed and agility, rather than vertical jump height, are the most significant discriminators between high-potential and average players.

### 6.4.2 Structural Equation Modelling (SEM)

As an extension of MANCOVA, *SEM* provides a more nuanced framework by incorporating latent variables and testing direct and indirect effects. We covered this earlier in the module.

### 6.4.3 Mixed-Design Approaches

Combining *MANOVA* or *MANCOVA* with *repeated measures designs* allows for the study of changes over time across multiple dependent variables. For example, a mixed-design MANOVA might evaluate how different dietary interventions influence performance metrics (speed, endurance) over a competitive season while controlling for baseline dietary habits.

### 6.4.4 Machine Learning Integration

As datasets in sports science grow, machine learning approaches can complement MANOVA and MANCOVA by handling complex, non-linear relationships. For instance, a *hybrid* model could predict injury risk by incorporating multivariate analyses alongside decision trees or neural networks.

## 7 References

- [1] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Hampshire, UK: Cengage Learning EMEA, 2019.
- [2] J. P. Stevens, *Applied Multivariate Statistics for the Social Sciences*, 5th ed. New York, NY: Routledge, 2009.