

Factorial ANOVA and MANOVA

Minsoo Kang¹ and Ying Jin²

¹ *Health and Human Performance, Middle Tennessee State University, Murfreesboro, TN, USA*

² *Quantitative Psychology, Middle Tennessee State University, Murfreesboro, TN, USA*

General Introduction

The ANOVA is a statistical method used to test mean differences among groups. In its simplest form, a one-way ANOVA with two groups functions the same as a *t*-test for independent samples. Unlike the *t*-test, however, more elaborate forms of ANOVA have the ability to test for mean differences among three or more groups of a single variable. The independent variable or variables in an ANOVA, also called factors or predictors, are categorical variables. Nominal and ordinal variables are common independent variables. The dependent variables are continuous (i.e., interval and ratio) variables.

The true advantage of the ANOVA, compared to *t*-test, comes from its ability to test for mean differences between two or more groups across multiple independent variables. When an ANOVA model has two independent variables, it is called a two-way ANOVA; when there are three independent variables, the ANOVA is called a three-way ANOVA; and so on. However, any ANOVA with two or more independent variables falls under the label of a factorial ANOVA. The multivariate analysis of variance (MANOVA) is an extension of the ANOVA that allows for multiple

dependent variables. The following sections will discuss the hypotheses and assumptions of a factorial ANOVA analysis with one dependent variable and also with multiple dependent variables (i.e., factorial MANOVA).

Hypothesis Testing

The factorial ANOVA and MANOVA test the main effects for each independent variable on a dependent variable, as well as any possible interactions between independent variables. A hypothesis test is required for each main effect and interaction effect.

A significant main effect indicates that the independent variable is a significant predictor of the dependent variable(s). A significant main effect in a factorial ANOVA indicates that at least one group in the independent variable differs on the dependent variable. In a factorial ANOVA, there is only one dependent variable for the independent variables to differ on. The null hypothesis for each main effect in a factorial MANOVA is that the dependent variables are the same across all conditions of that independent variable. The alternative hypothesis is that at least one group in the independent variable differs on at least one dependent variable.

For an interaction, the null hypothesis is that the effect of one independent variable on the dependent variable(s) is the same across each level of another independent variable. The alternative hypothesis is that the effect of an independent variable on the dependent variable(s) differs for the levels of another independent variable. For example, the benefits associated with being physically active (low, medium, and high) on health outcomes are likely to be different depending on an individual's body mass index (underweight, normal, overweight, and obese). To test this hypothesis, an interaction term between physical activity and body mass index groups is introduced and tested along with the main effects of these variables.

Alpha Level

Alpha level is related to the type I error rate, the chance that a researcher infers that there is a significant difference when the groups do not differ. The nominal alpha level is what the researcher sets as the chance of a type I error rate, usually 0.05 (or 5% chance). The empirical alpha level is the actual chance that a type I error occurs. Various approaches have been proposed to control the empirical alpha level in post hoc procedures. In a two-way factorial ANOVA, each main effects and interaction are considered a family and are compared to an alpha of 0.05. When the interaction is nonsignificant, the pairwise comparisons of the main effects are performed with a nominal alpha level of 0.05/comparison in order to control the familywise empirical alpha level close to 0.05. With a significant interaction, the nominal alpha level is split for each follow-up ANOVA ($0.05/\text{levels of other factor}$) and again for the number of pairwise comparisons ($[0.05/\text{levels of other factor}]/\text{comparison}$). When dealing with factorial MANOVAs, the same approach of splitting the alpha level is used until the analyses become univariate (one dependent variable) in nature. All univariate analyses retain the alpha level of the preceding multivariate analysis. This is called a protected F procedure. Even when the nominal alpha is controlled

at a familywise alpha of 0.05, the empirical alpha can differ drastically from the nominal alpha level if the following assumptions are not met.

Assumptions

Independence Assumption

The independence assumption means that the dependent variable scores of one participant are not affected by the dependent variable scores of another participant. ANOVAs are sensitive to violations of the independence assumption. Therefore, it is important to adhere to this assumption in two ways. First, each participant contributes only one score to the data; repeated assessments of the same participant require repeated measures ANOVA/MANOVA (see Chapter 2). Second, no outside influence can systematically affect the groups. For example, if two coaches are keeping track of the number of push-ups attainable by their players, and coach A accepts push-ups from the knees and coach B does not, the coaches have systematically influenced the measurements of their groups.

Normality Assumption

Normality refers to the distribution of the participants' scores on the dependent variable(s). The bell curve is the common conceptual image of univariate normality. Dealing with multiple dependent variables requires multivariate normality. Data that are multivariate normal are univariate normal as well. However, variables can be univariate normal without the data possessing a multivariate normal distribution. The effect of nonnormality on the ANOVA/MANOVA results is influenced by two main factors:

1. The extent of departure from normality (i.e., skewness and kurtosis)
2. Sample size (n)

The extent of the departure from normality diminishes in influence on test results with larger sample sizes. The influence due to departures from univariate normality on the ANOVA results is almost nonexistent when the sample size is above 30 participants per group (Myers, Well, & Lorch, 2010). The number of participants needed to reduce the effects of multivariate nonnormality relates to the number of dependent variables in the model. However, Seo, Kanda, and Fujikoshi (1995) demonstrated that the MANOVA is actually fairly robust to departures from multivariate normality with 10 participants per group. While there are statistical tests for normality (i.e., Kolmogorov–Smirnov test), these tests are more likely to indicate normality as sample size increases. As mentioned previously, the ANOVA and MANOVA are more robust to nonnormality as sample size increases. Often, a visual inspection of a histogram, Q–Q plot, or the skewness and kurtosis statistics can provide sufficient evidence for or against normality. Histograms can be produced with an added bell-shaped overlay to help assess normality. On a Q–Q plot, nonnormality is indicated by the dots straying from the diagonal line. Deviations from a bell shape are expected due to random error. Skewness and kurtosis can help describe just how far the data actually stray from normality. Skewness and kurtosis ratios can be calculated by

dividing the respective statistic by its standard error. Ratios greater than two or less than negative two are indicative of potential nonnormality.

Univariate: Homogeneity of Variance

The variance of the dependent variable should be the same for each group. The ANOVA is fairly robust to differences in variances as long as groups are of equal sample size (Maxwell & Delaney, 2004). In situations with small (<10 per group) or unequal sample sizes, it is best to utilize the Welch ANOVA (Myers et al., 2010) instead of the usual one-way ANOVA. The Welch ANOVA does not rely on the assumption of equal variance because it weights each group mean by its sample size. Levene's test provides a statistical test of the homogeneity of variance assumption. SPSS makes Levene's test (T2) available, but Nordstokke and Zumbo (2007) have shown that the empirical alpha level of Levene's test (T2) can be two to four times the nominal alpha level, resulting in a high chance of a type I error. Therefore, in situations with small or unequal sample sizes, it is recommended to not assume equal variance (Maxwell & Delaney, 2004).

Multivariate: Homogeneity of Variance–Covariance Matrix

Each group created by a unique combination of factor levels contains its own variance–covariance matrix for the dependent variables. The multivariate homogeneity of variance–covariance matrix assumption requires that this matrix is the same for each group. Box's M is available from SPSS for testing multivariate homogeneity of variance–covariance. With equal sample sizes, robustness can be expected regardless of Box's M significance (Tabachnick & Fidell, 2013).

Multicollinearity

When dealing with multiple dependent variables, each of these variables should be highly correlated with the independent variables but not highly correlated with the other dependent variables. If dependent variables are highly correlated with each other, multicollinearity can occur. This can drastically reduce the power of the MANOVA and potentially result in unstable solutions. In such cases, consider deleting one or more of the redundant dependent variables or use a principal component analysis (PCA) as a dimensionality reduction technique. The factors that are extracted by a PCA (assuming they are conceptually plausible) are completely uncorrelated and therefore address the issue of multicollinearity.

Further Considerations

A Priori Power Calculations

Power calculation should be done prior to data collection in order to identify the minimum number of participants required in order to achieve sufficient power. Without theory dictating an expected effect size, it is recommended to use a medium effect size (Maxwell & Delaney, 2004) with an alpha of 0.05 and an expected power of 0.80 (Tabachnick & Fidell, 2013). G*Power is a freeware program able to perform a priori power calculation.

Post Hoc Analyses Following a Significant Interaction

Significant interactions supersede main effects in a factorial ANOVA because comparisons of marginal means may no longer be appropriate. Significant interactions must be further analyzed in order to better understand the effects of the independent variables on the dependent variable(s). One approach to further analyze an interaction in any ANOVA with two or more factors is to perform simple effects ANOVAs. Simple effects tests analyze the effect of one or more factors at each level of another factor. The following SPSS syntax can be altered to fit any factorial ANOVA:

```
UNIANOVA Yvar
  BY Afactor Bfactor
  /CRITERIA ALPHA(.05/#LevelsBfactor)
  /EMMEANS = TABLES (Afactor* Bfactor) compare (Afactor) adj (SIDAK).
```

```
UNIANOVA Yvar
  BY Afactor Bfactor
  /CRITERIA ALPHA(.05/#LevelsAfactor)
  /EMMEANS = TABLES (Afactor* Bfactor) compare (Bfactor) adj (SIDAK).
```

where **Yvar** is the dependent variable and **Afactor** and **Bfactor** are the two independent variables. This syntax pools error terms and is therefore not appropriate in situations with unequal variance (Maxwell & Delaney, 2004). With unequal variance, the data must be manually split in order to calculate separate error terms for the simple effects ANOVAs and pairwise comparisons. The following SPSS syntax command will split the data by **Bfactor**:

```
SPLIT FILE SEPARATE BY Bfactor.
```

A Welch ANOVA may then be performed looking at the effect of **Afactor** for each level of **Bfactor**. Splitting the data again by **Afactor** allows for the effect of **Bfactor** to be analyzed for each level of **Afactor**.

Multivariate F Statistics

Four statistics are available for determining the significance of an independent variable's main or interaction effect in the MANOVA. Roy's largest root and Hotelling's trace are both liberal statistics and therefore overestimate actual significance. Wilks' lambda and Pillai's trace are much more conservative and less biased estimates of the significance. The values reported are the approximate *F* values associated with the multivariate statistic. Wilks' *F* is commonly reported. However, Pillai's *F*, being the most conservative, can help control the type I error rate with small or unequal sample sizes (Tabachnick & Fidell, 2013).

Univariate/Multivariate Effect Size

Effect sizes are important to report because as sample size increases, so does statistical power. With very large sample sizes, ANOVA will eventually produce a significant *F* test even though the differences between the groups are practically insignificant. Several effect sizes are available for the factorial ANOVA. Eta squared

(η^2) is easily calculated by hand as $[SS_{\text{Effect}}/SS_{\text{Total}}]$. Partial η^2 (η_p^2) is readily available from SPSS and is calculated as $[SS_{\text{Effect}}/(SS_{\text{Effect}} + SS_{\text{Error}})]$. In a one-way ANOVA, η^2 and η_p^2 are identical. Partial η^2 allows the effect size to only measure the unique variance accountable for by that independent variable. However, η^2 and η_p^2 are known to be overestimates, but η_p^2 is less biased than η^2 (Grissom & Kim, 2005). A more accurate and nearly unbiased measure of effect size is omega squared (ω^2) (Grissom & Kim, 2005). Partial η^2 and ω^2 are both considered appropriate. The formula for ω^2 is

$$\omega^2 = \frac{[SS_{\text{Effect}} - (df_{\text{Effect}} \times MS_{\text{Within}})]}{(SS_{\text{Total}} + MS_{\text{Within}})}$$

There is not a multivariate equivalent to ω^2 , but partial η_p^2 for multivariate effects is provided by SPSS.

Utility in Sport and Exercise Sciences

This section provides a discussion of the usefulness of the ANOVA model in the field of sport and exercise sciences. Examples of different types of independent variables are discussed briefly.

Treatment Conditions

Developing a new workout routine, contrasting coaching techniques, and comparing the health benefits of various types of physical activity all require the use of different treatment conditions. For example, in the development of a new workout routine, it may be useful to include commonly accepted workout routines similar to the one in development in order to compare and contrast the potential of the new routine. Not every group requires an actual treatment. The control group is a common addition to the treatment group(s). The control group allows the comparison to a comparable group of matched individuals who received no treatment.

Sisson et al. (2009) performed a study looking at fitness gains in women (45–75 years of age). In their study, they randomly assign women to one of three different exercise treatment conditions (4, 8, and 12 kcal/kg) with a fourth control group. The exercise routines were maintained for 6 months. While the control group was not of interest, it allowed Sisson et al. to have a comparable group of participants to control for other confounding effects (i.e., yearly weight fluctuations) that could influence the results of the treatment conditions.

Existing Conditions

It is not always possible for researchers to assign participants to certain groups. For instance, when analyzing injury rates between sports, sport membership is a preexisting condition. Working in schools also requires that researchers work with existing groups of students already assigned to different classes.

Beighle, Morgan, Le Masurier, and Pangrazi (2006) provide an example of an existing condition (grade level) in their study of children's physical activity in and out of school. Third, fourth, and fifth graders were monitored for step counts and amount of time spent being physically active during recess and outside of school. Beighle et al. (2006) found no significant difference between grade levels and no significant interaction between grade levels and gender. Boys were significantly more active than girls.

Individual Characteristics

Gender, ethnicity, and age groups are commonly included in ANOVAs in order to test for differences across individual characteristics. An interaction between new treatment condition (e.g., different workout routines) and an individual characteristic such as gender could indicate that a particular workout routine is more effective with men, while a different workout routine is more effective with women.

Seabra et al. (2013) used gender as an individual characteristic variable to study difference in the perception of physical activity and its enjoyment between boys and girls (8–10 years of age). Boys reported greater enjoyment of physical activity than girls.

Nyberg, Nordenfelt, Ekelund, and Marcus (2009) studied the differences between 6-, 7-, 8-, and 9-year-olds' physical activity levels for both boys and girls. Their study concluded that 9-year-olds are less physically active than 6-year-olds and boys are more physically active than girls.

Recent Usage

From 2010 through 2014, 12 peer-reviewed journals in the field of sport and exercise sciences were used to assess the frequency of the factorial ANOVA and MANOVA reporting. Appendix 1.5 provides the frequency of factorial ANOVA and MANOVA usage. The 12 peer-reviewed journals published a total of 8027 articles. Some form of a factorial ANOVA or MANOVA appeared in 519 of these articles. A one-way ANOVA or MANOVA was reported in 864 of these articles.

The Substantive Example

Kang and Brinthaup (2009) developed a school-based pedometer intervention program. Previous studies have suggested the use of group goals (every student is attempting to reach the same step count) or individual goals (5% of baseline increase) to increase walking. Extant research provided little direction on which goal one to choose and when. Kang and Brinthaup performed a 6-week study to analyze the effects of individual- and group-based step count goals. Groups differing in baseline physical activity level were included in the analysis as an independent variable to see if goal types were more effective for different physical activity levels. This section will use goal type and physical activity level as independent variables to analyze the differences in step goal attainment and postintervention (PI) step count.

Physical activity level for each participant was defined as their average step counts per day at baseline, broken down into low (<4800), medium, and high (>6300). Goal type was either individual- or group-based step count goal. The factorial ANOVA analysis will test if the number of days of goal attainment over the 6-week period differs by baseline physical activity level, goal type, or their interaction. The factorial MANOVA will keep the same model (for comparative purposes) and will introduce a second dependent variable, PI step counts.

Univariate: Factorial ANOVA

Dependent variable:

Number of days of goal attainment (hereafter referred to as goal attainment (days)): ratio data.

Independent variables and main effect hypotheses:

Baseline physical activity level: ordinal data with three levels [low ($n=31$), medium ($n=38$), and high ($n=30$)].

Null hypothesis: Physical activity level is not a significant predictor of goal attainment (days), or goal attainment (days) does not differ across physical activity levels.

Goal type: nominal data with two groups [(group-based ($n=57$) and individual-based ($n=42$)] step count goals.

Null hypothesis: Goal type is not a significant predictor of goal attainment (days), or goal attainment(days) does not differ across goal type.

Interaction effect hypothesis:

Physical activity level X goal type.

Null hypothesis: The effect of physical activity level on goal attainment (days) is the same across conditions of goal type.

Univariate Assumptions

Independence:

The data were collected from individual students who only contributed once to step count and goal attainment (days). Each wore his/her own pedometer, and no sharing of pedometers was reported. There is no evidence to suggest that the independence assumption was violated.

Normality:

Analysis of descriptive statistics indicates very low skewness or kurtosis in all the groups.

Homogeneity of variance:

Due to small and unequal sample sizes, the homogeneity of variance assumption is likely to have been violated. To maintain empirical alpha levels close to nominal alpha levels, the error terms will not be pooled. The data will be manually split in SPSS, and simple effects Welch ANOVAs and Games–Howell pairwise comparisons will be subsequently performed.

A priori power:

A priori power calculations were conducted using G*Power with the following parameters:

Effect size f : 0.25 (medium; Cohen, 1988)

α : 0.05

Power: 0.8

Numerator df : 2 ([3 levels of physical activity level – 1] * [2 levels of goal types – 1])

Number of groups: 6 (3 levels of physical activity level * 2 levels of goal type)

The resulting recommendation is a sample size of 158 that is not met by our sample size of 99 students. The example below is given for illustrative purposes. Adequate power is strongly advised for all studies.

Multivariate: Factorial MANOVA

Dependent variables:

Number of days of goal attainment (hereafter referred to as goal attainment (days)): ratio data.

Postintervention step count (hereafter referred to as PI step count): ratio data.

Independent variables and main effect hypotheses:

Physical activity level: same as before.

Null hypothesis: Physical activity level is not a significant predictor of goal attainment (days) and PI step count, *or* goal attainment (days) and PI step count do not differ across physical activity levels.

Goal type: same as before.

Null hypothesis: Goal type is not a significant predictor of goal attainment (days) and PI step count, *or* goal attainment (days) and PI step count do not differ across goal type.

Interaction effect hypothesis:

Physical activity level X goal type.

Null hypothesis: The effect of physical activity level on goal attainment (days) and PI step count is the same across conditions of goal type.

Multivariate Assumptions

Independence:

See univariate assumption.

Multivariate normality:

With more than 10 participants per group, the factorial MANOVA is fairly robust to any divergence from multivariate normality. In addition, no indication of univariate nonnormality, while not sufficient to demonstrate multivariate normality, also suggests that the normality assumption is valid for all post hoc analyses.

Homogeneity of variance–covariance:

Due to small and unequal sample sizes, the homogeneity of variance–covariance assumption is likely to have been violated. To maintain empirical alpha levels close

to nominal alpha levels, the error terms will not be pooled. The data will be manually split in SPSS, and subsequently simple effects Welch ANOVAs and Games–Howell pairwise comparisons will be performed. Pillai’s F will be reported in order to keep in control type I error rate.

A priori power:

A priori power calculations were conducted using G*Power with the following parameters:

Effect size $f^2(V)$: 0.15 (medium; Cohen, 1988)

α err prob: 0.05

Power (1- β err prob): 0.8

Number of groups: 6 (3 levels of physical activity level * 2 levels of goal type)

Number of predictors: 2 (# of independent variables)

Response Variables: 2 (# of dependent variables)

The resulting recommendation is a sample size of 43 that is met by our sample of 99 students.

The Synergy

This section will illustrate the analysis plan for a factorial ANOVA and a factorial MANOVA and the reporting of the output in a way that is compatible with the American Psychological Association (APA) publication manual. We will also highlight what is most important and necessary to report. The SPSS syntax and data can be found in the Appendices 1.1 and 1.3.

Factorial ANOVA Analysis Plan

The full model was analyzed for the main effects and any possible interactions (see Figure 1.1 for the ANOVA analysis plan). In this example, the main effects of physical activity and goal type and the interaction of the two independent variables were explored (see Appendix 1.2 for the abbreviated output).

This example contains a significant interaction. With equal variance, the EMMEANS command could have been used in order to produce the necessary post hoc analyses. Given a significant interaction with unequal variance, the data were split by one of the variables involved in the interaction, analyzed, split by the other variable, and analyzed again. For this example, the data were split first by goal type, and the effect of physical activity level was analyzed and reported for each goal type ($\alpha = 0.05/2$ goal types = 0.025). Physical activity level is comprised of three levels and thus requires follow-up comparisons. Given the small and unequal sample sizes, Games–Howell pairwise comparisons were analyzed. The Games–Howell procedure is commonly used for situations where equal variance cannot be assumed, as it is the same as Tukey’s HSD adjusted for each pairwise comparison.

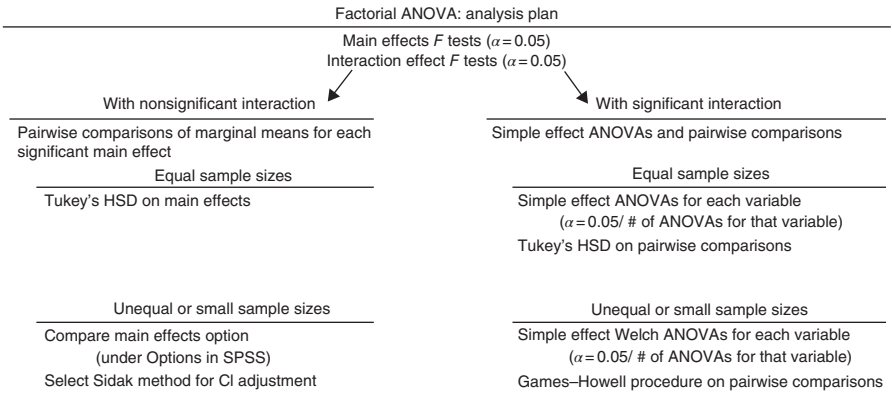


Figure 1.1 Factorial ANOVA analysis plan.

Next, the data were split by physical activity level, and the effect of goal type was analyzed and reported for each level of physical activity level ($\alpha = 0.05/3$ levels of physical activity = 0.0167). Given that goal type is a dichotomous variable, a follow-up pairwise comparison is redundant with the simple effects ANOVA.

If the interaction was not significant, comparisons of the marginal means would be reported for each significant main effect with Tukey's HSD or the Sidak adjustment for small or unequal sample (Games–Howell is unavailable for comparisons of marginal means). Tukey's HSD and Sidak assume equal variance, while the Sidak is slightly more conservative.

Example of a Write-Up Compatible with the APA Publication Manual

A two-way ANOVA was used to evaluate the number of days that each student attained their goal based on their goal type and physical activity level (see Table 1.1 for descriptive statistics). Physical activity level and goal type were significant predictors of the number of days of goal attainment, $F(2, 93) = 17.67$, mean square error (MSE) = 33.33, $p < 0.001$, $\eta_p^2 = 0.275$, and $F(1, 93) = 6.85$, $p = 0.010$, $\eta_p^2 = 0.069$, respectively. There was a significant interaction between goal type and physical activity level, $F(2, 93) = 10.25$, $p < 0.001$, $\eta_p^2 = 0.181$ (Figure 1.3). As a result, simple effects ANOVAs were conducted to probe the interaction effect.

Welch ANOVAs and Games–Howell pairwise comparisons were used to conduct simple effects tests for predicting goal attainment based on physical activity level for each goal type (see Table 1.2 for pairwise comparisons). The alpha for each ANOVA was 0.025 because there were two different groups in goal type. The results indicated that physical activity level was a significant predictor of goal attainment for students in the group goal program, $F(2, 35.04) = 38.17$, MSE = 34.37, $p < 0.001$, $\eta_p^2 = 0.529$. High activity level student attained significantly more goals than middle activity level students who also attained significantly more goals than low activity level students.

Table 1.1 Descriptive statistics for goal attainment (days) and postintervention (PI) step count.

	Goal attainment (days)		PI step count		<i>n</i>
	<i>M</i>	SD	<i>M</i>	SD	
<i>Group goal</i>					
Physical activity level					
Low	6.22	4.88	5467.06	1029.44	18
Middle	15.57	6.74	6121.00	1363.23	23
High	21.69	5.49	7883.25	2084.39	16
<i>Individual goal</i>					
Physical activity level					
Low	16.38	3.10	4493.77	1402.98	13
Middle	17.93	7.23	7065.33	2486.28	15
High	18.43	5.53	7970.71	1720.71	14

Table 1.2 Factorial ANOVA: Games–Howell comparisons of days of goal attainment (days).

(I)	(J)	Mean difference	97.5% CI	
		(I-J)	Lower limit	Upper limit
<i>Comparisons for physical activity level</i>				
Subgroup=group goal				
Physical activity level				
Low	Middle	-9.34*	-14.31	-4.38
Low	High	-15.47*	-20.42	-10.51
Middle	High	-6.12*	-11.51	-0.74
Subgroup=individual goal				
Physical activity level				
Low	Middle	-1.55	-7.44	4.34
Low	High	-2.04	-6.92	2.83
Middle	High	-0.50	-7.16	6.16

*Significant at a familywise alpha of 0.05 (0.025 per simple effects).

Physical activity level was not a significant predictor of goal attainment for students in the individual goal program, $F(2, 24.08) = 0.82$, $MSE = 31.88$, $p = 0.450$, $\eta_p^2 = 0.024$. The pairwise comparisons were nonsignificant.

Simple effects Welch ANOVAs were used to predict the number of days of goal attainment based on the goal type for each level of physical activity. An alpha of 0.0167 was used to determine significance for the simple effects ANOVAs. No pairwise comparisons were used as there are only two groups within goal type. Low physical activity students in the individual goal program attained their goal more often

than the group goal program, $F(1, 28.63)=50.10$, $MSE=17.94$, $p<0.001$, $\eta_p^2=0.600$. However, goal type was not a significant predictor of goal attainment for middle and high physical activity level students, $F(1, 28.55)=1.03$, $MSE=48.07$, $p=0.319$, $\eta_p^2=0.029$, and $F(1, 27.41)=2.613$, $MSE=30.32$, $p=0.117$, $\eta_p^2=0.085$, respectively.

Factorial MANOVA Analysis Plan

As with the ANOVA analysis plan, the full model was analyzed in order to inform the direction of the post hoc analyses. Wilks' lambda is commonly reported and considered to be sufficient at controlling type I error rates. Pillai's is the most conservative and should be used with small or unequal sample sizes in order to control for type I error rate. In this example, Pillai's F statistic will be reported for the main effects of physical activity level and goal type and their interaction.

Post hoc analyses will be conducted using the protected F procedure. The protected F test is a commonly used procedure to further analyze the effects of the independent variables on the dependent variables.

The same variables from the factorial ANOVA example are included and will provide the same significant interaction. This is intentional in order to highlight the process of simple effects MANOVAs with a protected F procedure. The data are split first by goal type, and the effect of physical activity level is analyzed in simple effects MANOVAs for each goal type ($\alpha=0.05/2$ goal types= 0.025). Where physical activity is a significant predictor, follow-up Welch ANOVAs will be reported for physical activity level predicting goal attainment (days) and PI step count ($\alpha=0.025$). Games–Howell pairwise comparisons will be used because physical activity level is comprised of three levels.

Next, the data are split by physical activity level, and the effect of goal type is analyzed in simple effects MANOVAs for each level of physical activity level ($\alpha=0.05/3$ level of physical activity= 0.0167). Where goal type is a significant predictor, follow-up Welch ANOVAs, due to unequal variance, will be reported for goal type predicting goal attainment (days) and PI step count ($\alpha=0.0167$).

With a nonsignificant interaction, the independent variables are analyzed in a factorial ANOVA plan for each dependent variable (see Figure 1.2 for the MANOVA analysis plan). This reduces the MANOVA down to a series of factorial ANOVAs, each with its own analysis plan.

Example of a Write-Up Compatible with the APA Publication Manual

A two-way MANOVA was used to evaluate the number of days that each student attained their goals and their PI step count based on their physical activity level and goal type. An alpha of 0.05 was used. Physical activity level and goal type were both significant predictors, Pillai's $F(4, 186)=9.49$, $p<0.001$, $\eta_p^2=0.169$, and Pillai's $F(2, 92)=8.05$, $p=0.001$, $\eta_p^2=0.149$. There was a significant interaction between physical activity level and goal type, Pillai's $F(4, 186)=15.39$, $p<0.001$, $\eta_p^2=0.249$ (see Figures 1.3 and 1.4). As a result, simple effects MANOVAs were conducted.

Simple effects MANOVAs were conducted to predict the number of days that each student attained their goal and their PI step count based on the physical

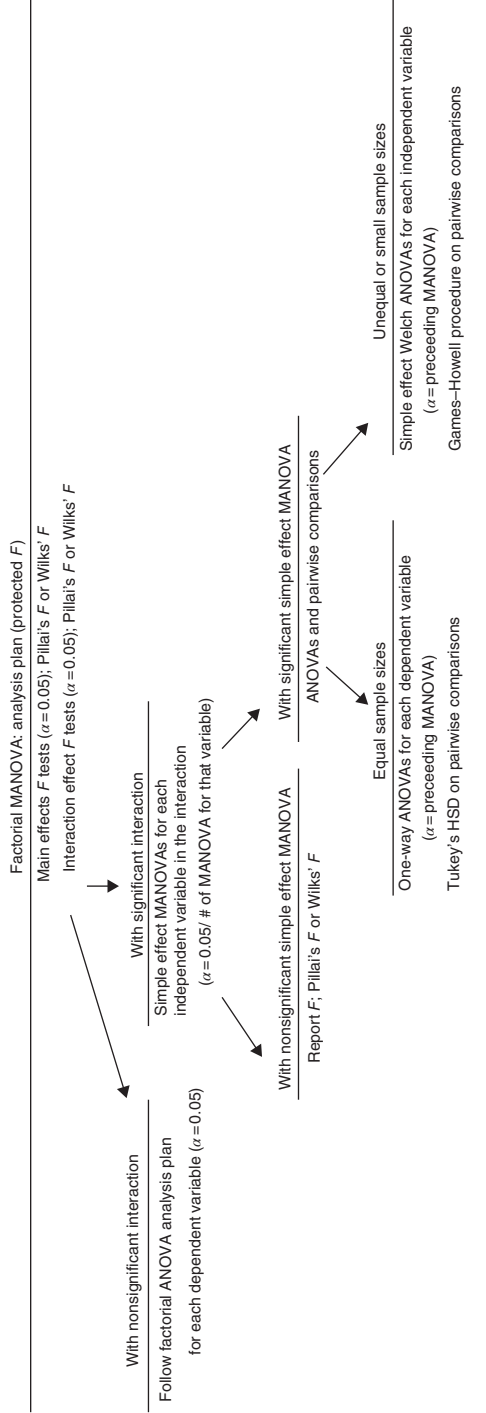


Figure 1.2 Factorial MANOVA analysis plan (protected F).

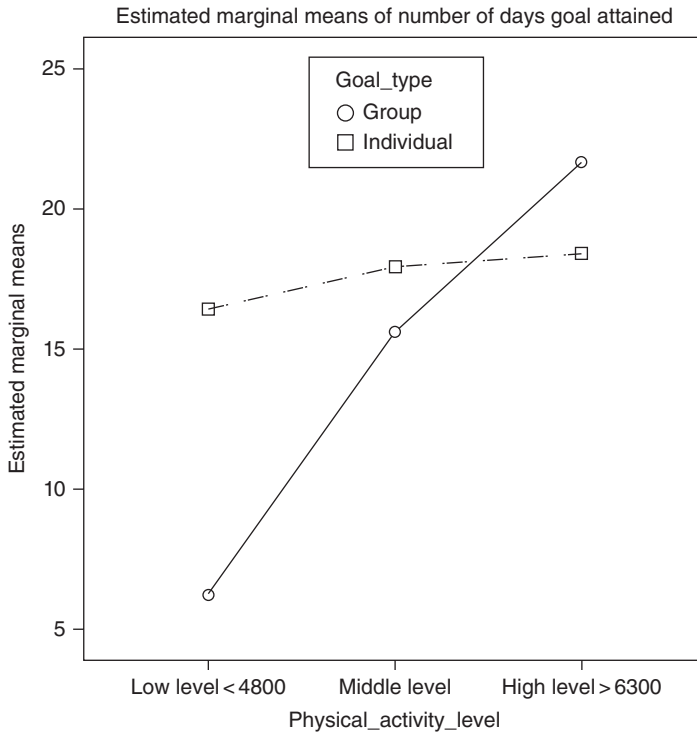


Figure 1.3 Marginal means plot of goal attainment (days) (a clear interaction can be visualized by the differences in slopes between individual and group goal types).

activity level for each goal type. An alpha of 0.025 was used to determine significance. For the group goal program, physical activity level was a significant predictor, Pillai's $F(4, 108) = 13.81, p < 0.001, \eta_p^2 = 0.338$. Follow-up Welch ANOVAs, with an alpha of 0.025, and Games–Howell pairwise comparisons were conducted (see Table 1.3 for pairwise comparisons). Physical activity level was a significant predictor of goal attainment, $F(2, 35.04) = 38.17, p < 0.001, \eta_p^2 = 0.529$. High activity level student attained more goals than middle activity level students who also attained significantly more goals than low activity level students. Physical activity level was also a significant predictor of PI step count, $F(2, 31.63) = 8.84, p = 0.001, \eta_p^2 = 0.298$. High activity level students had significantly higher PI step count than middle and low activity level students; the latter two groups did not differ significantly from each other. For the individual goal program, physical activity level was a significant predictor, Pillai's $F(4, 78) = 7.41, p < 0.001, \eta_p^2 = 0.275$. Follow-up Welch ANOVAs, with an alpha of 0.025, indicated that physical activity level was not a significant predictor of goal attainment, $F(2, 24.08) = 0.82, p = 0.450, \eta_p^2 = 0.024$, but was a significant predictor of PI step count, $F(2, 25.61) = 17.46, p < 0.001, \eta_p^2 = 0.370$. Low activity level students maintained lower step count compared to middle and high activity level students who did not differ in step count.

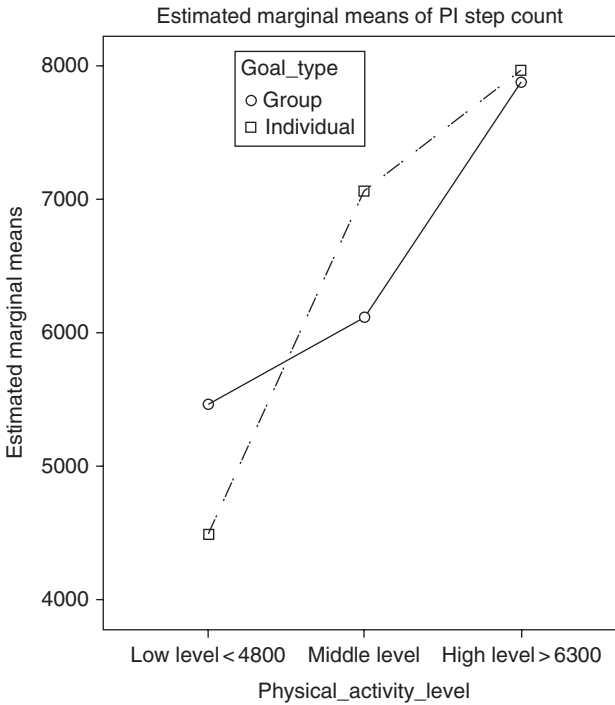


Figure 1.4 Marginal means plot of postintervention step count (a clear interaction can be seen between low- and middle-level physical activity and goal type).

Simple effects MANOVAs were conducted to predict the number of days that each student attained their goal and their PI step count based on the goal type for each physical activity level. An alpha of 0.0167 was used to determine significance for the simple effects MANOVAs. Goal type was a significant predictor for low activity level students, *Pillai's* $F(2, 28)=55.24$, $p<0.001$, $\eta_p^2=0.798$. Follow-up Welch ANOVAs, alpha=0.0167, indicated that low activity level students in the individual goal program attained their goal more often than those in the group goal program, but no difference was found in PI step count, $F(1, 28.63)=50.10$, $p<0.001$, $\eta_p^2=0.600$, and $F(1, 20.91)=4.51$, $p=0.046$, $\eta_p^2=0.147$ (recall that an alpha of 0.167 was used to establish significance). Goal type was not a significant predictor for middle and high activity level students, *Pillai's* $F(2, 35)=1.19$, $p=0.317$, $\eta_p^2=0.064$, and *Pillai's* $F(2, 27)=3.33$, $p=0.051$, $\eta_p^2=0.198$, respectively.

Summary

The factorial ANOVA and MANOVA are natural extensions of the one-way ANOVA. The strength of these tests is that multiple independent variables can be analyzed simultaneously, testing their main and interaction effects. The example in this chapter reported an interaction within the data and demonstrated how to further probe it.

Table 1.3 Factorial MANOVA: Games–Howell comparisons of days of goal attainment (days) and PI step count.

(I)	(J)	Mean difference	97.5% CI	
		(I–J)	Lower limit	Upper limit
<i>Comparisons for physical activity level on goal attainment (days)</i>				
Subgroup = group goal				
Physical activity level				
Low	Middle	–9.34*	–14.31	–4.38
Low	High	–15.47*	–20.42	–10.51
Middle	High	–6.12*	–11.51	–0.74
Subgroup = individual goal				
Physical activity level				
Low	Middle	–1.55	–7.44	4.34
Low	High	–2.04	–6.92	2.83
Middle	High	–0.50	–7.16	6.16
<i>Comparisons for physical activity level on PI step count</i>				
Subgroup = group goal				
Physical activity level				
Low	Middle	–653.94	–1674.70	366.81
Low	High	–2416.19*	–4051.36	–781.03
Middle	High	–1762.25*	–3434.84	–89.66
Subgroup = individual goal				
Physical activity level				
Low	Middle	–2571.56*	–4695.94	–447.19
Low	High	–3476.95*	–5169.98	–1783.91
Middle	High	–905.38	–3122.20	1311.44

*Significant at a familywise alpha of 0.05 (0.025 per simple effects).

This chapter covered the theoretical framework of the factorial analyses starting with hypothesis testing. The hypotheses are laid out prior to analysis and help to drive the interpretation of the results. The assumptions of the ANOVA are necessary in order to maintain the power of the analysis. Keeping high numbers in terms of group membership (where possible) and minimizing attrition rates can help to keep sample sizes large and equal. Large and equal sample sizes guard against unequal variance and ensure that researchers have enough degrees of freedom for adequate power. The independence assumption should be considered during the design of the study. If the independence assumption is violated, it can have a detrimental effect on the type I error rate within the analysis procedures.

The analysis plan laid out within this chapter is easy to generalize and adapt to factorial ANOVAs with more than 2 predictors. Significant interactions can occur with three- and four-way interactions. The highest interactions term should be analyzed first through splitting the data by each variable involved. The simple effects tests provide a strong and easily interpreted approach to isolating the way that

independent variables differ in their effect on the dependent variable across levels of other independent variables. Descriptive statistics (M , SD , n) are best presented in a table. However, if only a few variables are analyzed, the descriptive statistics can be presented in the write-up when the variables are first mentioned. If model summary statistics (F , MSE , R^2 , and p) are not presented within the Results section, they should be included in a model summary table. Pairwise comparisons of either the marginal means or the simple effects pairwise comparisons (e.g., Tukey's or Games–Howell) should be presented in a table but still discussed in the write-up.

Acknowledgment

The authors would like to thank Hershel Eason from Middle Tennessee State University for his contribution in the preparation of this chapter.

References

- Beighle, A., Morgan, C. F., Le Masurier, G., & Pangrazi, R. P. (2006). Children's physical activity during recess and outside of school. *Journal of School Health*, 76(10), 516–520.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kang, M., & Brinthaup, T. (2009). Effects of group- and individual-based step goals on children's physical activity levels in school. *Pediatric Exercise Science*, 21, 148–158.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Myers, J. L., Well, A. D., & Lorch, R. F., Jr. (2010). *Research design and statistical analysis* (3rd ed.). New York, NY: Routledge.
- Nordstokke, D. W., & Zumbo, B. D. (2007). A cautionary tale about Levene's tests for equal variances. *Journal of Educational Research and Policy Studies*, 7, 1–14.
- Nyberg, G. A., Nordenfelt, A. M., Ekelund, U., & Marcus, C. (2009). Physical activity patterns measured by accelerometry in 6- to 10-yr-old children. *Medicine and Science in Sports and Exercise*, 41(10), 1842–1848.
- Seabra, A., Mendonça, D., Maia, J., Welk, G., Brustad, R., Fonseca, A. M., & Seabra, A. F. (2013). Gender, weight status and socioeconomic differences in psychosocial correlates of physical activity in schoolchildren. *Journal of Science and Medicine in Sport*, 16(4), 320–326.
- Seo, T., Kanda, T., & Fujikoshi, Y. (1995). The effects of nonnormality on tests for dimensionality in canonical correlation and MANOVA models. *Journal of Multivariate Analysis*, 52, 325–337.
- Sisson, S. B., Katzmarzyk, P. T., Earnest, C. P., Bouchard, C., Blair, S. N., & Church, T. S. (2009). Volume of exercise and fitness nonresponse in sedentary, postmenopausal women. *Medicine and Science in Sports and Exercise*, 41(3), 539–545.
- Tabachnick, B. C., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.