UNDERSTANDING THE TWO-WAY ANOVA

We have seen how the one-way ANOVA can be used to compare two or more sample means in studies involving a single independent variable. This can be extended to two independent variables (still maintaining a single dependent variable, which falls under the general heading of univariate statistics). It should come as no surprise that the kind of ANOVA to be considered here is referred to as a two-way ANOVA. A minimum of four \( \bar{X}s \) are involved in any two-way ANOVA (i.e., two independent variables with a minimum of two levels each).

Like any one-way ANOVA, a two-way ANOVA focuses on group means. Because it is an inferential technique, any two-way ANOVA is actually concerned with the set of \( \mu \) values that correspond to the sample means that are computed from study’s data. Statistically significant findings deal with means. If the phrase on average or some similar wording does not appear in the research report your readers may make the literal interpretation of that all of the individuals in one group outperformed those in the comparison group(s). Statistical assumptions may need to be tested, and the research questions will dictate whether planned and/or post hoc comparisons are used in conjunction with (or in lieu of) the two-way ANOVA.

The two-way ANOVA has several variations of its name. For example, given that a factor is an independent variable, we can call it a two-way factorial design or a two-factor ANOVA. Another alternative method of labeling this design is in terms of the number of levels of each factor. For example, if a study had two levels of the first independent variable and five levels of the second independent variable, this would be referred to as a \( 2 \times 5 \) factorial or a \( 2 \times 5 \) ANOVA.

Factorial ANOVA is used to address research questions that focus on the difference in the means of one dependent variable when there are two or more independent variables. For example, consider the following research question: is there a difference in the fitness levels of men and women who fall into one of three different age groups? In this case we have one dependent variable (i.e., fitness) and two independent variables (i.e., gender and age). The first of these, gender, has two levels (i.e., Men and Women), and the second, age, has three levels. The appropriate statistical test to address this question is known as a \( 2 \times 3 \) factorial ANOVA, where the number 2 and 3 refer to the number of levels in the two respective independent variables.

THE VARIABLES IN THE TWO-WAY ANOVA

There are two kinds of variables in any ANOVA

1. **INDEPENDENT VARIABLE**
   - A variable that is controlled or manipulated by the researcher. A categorical (discrete) variable used to form the groupings of observations.
   - There are two types of independent variables: active and attribute.
     - If the independent variable is an active variable then we manipulate the values of the variable to study its affect on another variable. For example, anxiety level is an active independent variable.
     - An attribute independent variable is a variable where we do not alter the variable during the study. For example, we might want to study the effect of age on weight. We cannot change a person’s age, but we can study people of different ages and weights.
A two-way ANOVA always involves two independent variables. Each independent variable, or factor, is made up of, or defined by, two or more elements called levels. Sometimes factors are called independent variables and sometimes they are called main effects. When looked at simultaneously, the levels of the first factor and the levels of the second factor create the conditions of the study to be compared. Each of these conditions is referred to as a cell.

2. DEPENDENT VARIABLE

- The continuous variable that is, or is presumed to be, the result (outcome) of the manipulation of the independent variable(s).

In the two-way (two-factor) ANOVA, there are two independent variables (factors) and a single dependent variable.

- Changes in the dependent variable are, or are presumed to be, the result of changes in the independent variables.
- Generally, the first independent variable is the variable in the rows (typically identified as J), and the second variable is in the columns (typically identified as K).

We will use a new symbol called the dot subscript to denote the means of the rows and columns.

- For example, $\bar{X}_1$ is the mean of all observations in the first row of the data matrix averaged across all columns.
- $\bar{X}_j$ is the mean of all observations in the first column of the data matrix averaged across the rows.
- The mean of the observations in the first cell of the matrix corresponding to the first row and the first column is denoted $\bar{X}_{11}$.
- In general, $\bar{X}_{jk}$ is the cell mean for the $j$th row and the $k$th column, $\bar{X}_j$ is the row mean for the $j$th row, and $\bar{X}_k$ is the column mean for the $k$th column.

ADVANTAGES OF A FACTORIAL DESIGN

One reason that we do not use multiple one-way ANOVAs instead of a single two-way ANOVA is the same as the reason for carrying out a one-way ANOVA rather than using multiple $t$-tests, for to do so would involve analyzing in part the same data twice and therefore there may be an increased risk of committing a Type I error. However, there is one additional and very important reason for carrying out a factorial ANOVA in preference to using multiple one-way ANOVA tests. This reason is that carrying out multiple one-way ANOVAs would not enable the researcher to test for any interactions between the independent variables. Therefore a factorial ANOVA is carried out in preference to multiple one-way ANOVAs to avoid any increased risk in committing a Type I error and to enable both main and interaction effects to be tested.

- A researcher may decide to study the effects of $K$ levels of a single independent variable (or factor) on a dependent variable. Such a study assumes that effects of other potentially important variables are the same over the $K$ levels of this independent variable.
• An alternative is to investigate the effects of one independent variable on the dependent variable, in conjunction with one or more additional independent variables. This would be called a factorial design.
• There are several advantages to using a factorial design:
  1. One is efficiency (or economy)…
     • With simultaneous analysis of the two independent variables, we are in essence carrying out two separate research studies concurrently.
     • In addition to investigating how different levels of the two independent variables affect the dependent variable, we can test whether levels of one independent variable affect the dependent variable in the same way across the levels of the second independent variable.
     • If the effect is not the same, we say there is an interaction between the two independent variables.
     • Thus, with the two-factor design, we can study the effects of the individual independent variables, called the main effects, as well as the interaction effect.
     • Since we are going to average the effects of one variable across the levels of the other variable, a two-variable factorial will require fewer participants than would two one-ways for the same degree of power.
  2. Another advantage is control over a second variable by including it in the design as an independent variable.
  3. A third advantage of factorial designs is that they allow greater generalizability of the results. Factorial designs allow for a much broader interpretation of the results, and at the same time give us the ability to say something meaningful about the results for each of the independent variables separately.
  4. The fourth (and perhaps the most important) advantage of factorial designs is that it is possible to investigate the interaction of two or more independent variables.
     • Because of the real world of research, the effect of a single independent variable is rarely unaffected by one or more other independent variables, the study of interaction among the independent variables may be the more important objective of an investigation. If only single-factor studies were conducted, the study of interaction among independent variables would be impossible.
     • In addition to investigating how different levels of the two independent variables affect the dependent variable, we can test whether levels of one independent variable affect the dependent variable in the same way across the levels of the second independent variable. If the effect is not the same, we say that there is an interaction between the two independent variables.
**Models in the Two-way ANOVA**

In an ANOVA, there are two specific types of models that describe how we choose the levels of our independent variable. We can obtain the levels of the treatment (independent) variable in at least two different ways: We could, and most often do, deliberately select them or we could sample them at random. The way in which the levels are derived has important implications for the generalization we might draw from our study.

If the levels of an independent variable (factor) were selected by the researcher because they were of particular interest and/or were all possible levels, it is a **fixed-model** (fixed-factor or effect). In other words, the levels did not constitute random samples from some larger population of levels. The treatment levels are deliberately selected and will remain constant from one replication to another. Generalization of such a model can be made only to the levels tested.

Although most designs used in behavioral science research are fixed, there is another model we can use. If the levels of an independent variable (factor) are randomly selected from a larger population of levels, that variable is called a **random-model** (random-factor or effect). The treatment levels are randomly selected and if we replicated the study we would again choose the levels randomly and would most likely have a whole new set of levels. Results can be generalized to the population levels from which the levels of the independent variable were randomly selected.

Applicable to the factorial design, if the levels of one of the independent variables are fixed and the levels of the second independent variable are random, we have a **mixed-effects model**. In a mixed-model, the results relative to the random effects can be generalized to the population of levels from which the levels were selected for the investigation; the results relative to the fixed effect can be generalized to the specific levels selected.

**Hypotheses for the Two-way ANOVA**

- The null hypothesis for the $J$ row population means is
  - $H_0: \mu_1 = \mu_2 = \ldots = \mu_J$.
  - That is, there is no difference among the $J$ row means.
  - The alternative hypothesis would be that the $J$ row $\mu$s are *not* all equal to each other. That is, at least one pair or combination of $J$ row means differs.

- The null hypothesis for the $K$ column population means is
  - $H_0: \mu_{1.} = \mu_{2.} = \ldots = \mu_{K.}$
  - That is, there is no difference among the $K$ column means.
  - The alternative hypothesis would be that the $K$ column $\mu$s are *not* all equal to each other. That is, at least one pair or combination of $K$ column means differs.
The null hypothesis for the $JK$ interaction is

- $H_0$: all $(\mu_{JK} - \mu_J - \mu_{.K} + \mu) = 0$
  - That is, there is no difference in the $JK$ cell means that cannot be explained by the differences among the row means, the column means, or both.
  - The alternative hypothesis would be that the pattern of differences among the cell $\mu$s in the first column (or the first row) fails to describe accurately the pattern of differences among the cell $\mu$s in at least one other column (row). That is, there are differences among the cell population means that cannot be attributable to the main effects. In other words, there is an interaction between the two independent variables.

**Assumptions Underlying the Two-way ANOVA**

A two-way ANOVA is a parametric test, and as such, it shares similar assumptions to all other parametric tests. Recall that parametric statistics are used with data that have certain characteristics – approximate a normal distribution – and are used to test hypotheses about the population. In addition to those listed below, the dependent variable should be measured on an interval (continuous) scale and the factors (independent variables) should be measured on a categorical or discrete scale.

The importance of the assumptions is in the use of the $F$ distribution as the appropriate sampling distribution for testing the null hypotheses. Similar to the assumptions underlying the one-way ANOVA, the assumptions underlying the two-way (factorial) ANOVA are:

1. The samples are independent, random samples from defined populations. This is commonly referred to as the **assumption of independence**.
   - Independence and randomness are methodological concerns; they are dealt with (or should be dealt with) when a study is set up, when data are collected, and when results are generalized beyond the participants and conditions of the researcher’s investigation. Although the independence and randomness assumption can ruin a study if they are violated, there is no way to use the study’s sample data to test the validity of these prerequisite conditions. This assumption is therefore “tested” by examining the research design of the study.

2. The scores on the dependent variable are normally distributed in the population. This is commonly referred to as the **assumption of normality**.
   - This assumption can be tested in certain circumstances and should be tested using such procedures as a Shapiro-Wilks.

3. The population variances in all cells of the factorial design are equal. This is commonly referred to as the assumption of **homogeneity of variance**.
   - This assumption can be tested in certain circumstances and should be tested using such procedures as a Levene’s test of Homogeneity.

The consequences of violating one or more of these assumptions are similar to that of the one-way ANOVA. As is the case for one-way ANOVA, the two-way ANOVA is robust with respect to the violations of the assumptions, particularly when there are large and equal numbers of observations in each cell of the factorial.
Researchers have several options when it becomes apparent that their data sets are characterized by extreme non-normality or heterogeneity of variance. One option is to apply a data transformation before testing any null hypotheses involving main effects or cell means. Different kinds of transformations are available because non-normality or heterogeneity of variance can exist in different forms. It is the researcher’s job to choose an appropriate transformation that accomplishes the desired objective of bringing the data into greater agreement with the normality and equal variance assumptions.

Ideally, for an interaction to be present the two independent variables should not significantly correlate with each other, however, each independent variable should correlate significantly with the dependent variable.

**The Meaning of Main Effects**

With the two-way ANOVA, there are two main effects (i.e., one for each of the independent variables or factors). Recall that we refer to the first independent variable as the $J$ row and the second independent variable as the $K$ column. For the $J$ (row) **main effect**... the row means are averaged across the $K$ columns. For the $K$ (column) **main effect**... the column means are averaged across the $J$ rows. We can look at the main effects in relation to research questions for the study. In any two-way ANOVA, the first research question asks whether there is a statistically significant main effect for the factor that corresponds to the rows of the two-dimensional picture of the study. That is, the first research question is asking whether the main effect means associated with the first factor are further apart from each other than would be expected by chance alone. There are as many such means as there are levels of the first factor. The second research question in any two-way ANOVA asks whether there is a statistically significant main effect for the factor that corresponds to the columns of the two-dimensional picture of the study. We would consider the column main effect to be statistically significant if the main effect means for the second factor turn out to be further apart from each other than would be expected by chance alone.

If there are only two levels of either $J$ or $K$, and they are found to be significant, formal post hoc procedures are not necessary. Simply examine the group means for interpretation of significant differences. If there are three or more levels of either $J$ or $K$, and they are found to be significant, formal post hoc procedures are required to determine which pairs or combinations of means differ. The choice of post hoc (multiple-comparison) procedures is dependent upon several factors – with the key consideration as to whether the assumption of homogeneity of variance has been met (e.g., Tukey) or violated (e.g., Games-Howell).

**The Meaning of Interaction**

An interaction between the two factors is present in a two-way ANOVA when the effect of the levels of one factor is not the same across the levels of the other factor (or vice versa). For the two-way ANOVA, a third research question exists that asks whether there is a statistically significant interaction between the two factors involved in the study. The interaction deals with the cell means, not main effect means. An interaction exists to the extent that the difference between the levels of the first factor changes when we move from level to level of the second factor. There will always be at least four means involved in the interaction of any $2 \times 2$ ANOVA, six means involved in any $2 \times 3$ ANOVA, and so on.
The indication of a significant interaction is an examination of the $F$ ratio for the interaction – comparing the probability value to the alpha level. If $p \leq \alpha$, we would reject the null hypothesis and conclude that there is a significant interaction between the two independent variables. Conversely, if $p > \alpha$, we would retain the null hypothesis and conclude that there is not a significant interaction between the two independent variables.

One procedure for examining an interaction is to plot the cell means. The scale of the dependent variable is placed on the vertical (Y) axis; the level of one of the independent variables is placed on the horizontal (X) axis, and the lines represent the second independent variable. The decision about which variable to place on the X axis is arbitrary. The decision is typically based on the focus of the study. That is, are we concerned primarily (or exclusively) with the row (or column) effect (see Simple Main Effects). Regardless of which variable is placed on this axis, the interpretation of the interaction is, for all practical purposes, the same.

If there is no significant interaction between the two independent variables, the lines connecting the cell means will be parallel or nearly parallel (within sampling fluctuation). The plot of a significant interaction (the lines are not parallel) can have many patterns. If the levels (lines) of one independent variable never cross at any level of the other independent variable – this pattern is called an ordinal interaction – in part because there is a consistent order to the levels. That is, when one group is consistently above the other group, we have an ordinal interaction. If the levels (lines) of one independent variable cross at any level of the other independent variable – this pattern is called a disordinal interaction. That is, an interaction in which group differences reverse their sign at some level of the other variable is referred to as a disordinal interaction. What we need to keep in mind is that our interpretation is based on the data from the study. We cannot speculate as to what may (or may not) occur beyond the given data points. For example, if we look at the ordinal interaction plot, it would appear that if the lines were to continue, they would eventually cross. This is an interpretation that we cannot make, as we do not have data beyond those data points shown in the plots.
### Summary Table for the Two-way ANOVA

**Summary ANOVA: The General Case for Two-way ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degrees of Freedom (df)</th>
<th>Variance Estimate (Mean Square, MS)</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>(SS_J)</td>
<td>(J - 1)</td>
<td>(MS_J = \frac{SS_J}{J - 1}) (\frac{MS_J}{MS_W})</td>
<td></td>
</tr>
<tr>
<td>Columns</td>
<td>(SS_K)</td>
<td>(K - 1)</td>
<td>(MS_K = \frac{SS_K}{K - 1}) (\frac{MS_K}{MS_W})</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>(SS_{JK})</td>
<td>((J - 1)(K - 1))</td>
<td>(MS_{JK} = \frac{SS_{JK}}{(J - 1)(K - 1)}) (\frac{MS_{JK}}{MS_W})</td>
<td></td>
</tr>
<tr>
<td>Within-Cells</td>
<td>(SS_W)</td>
<td>(JK(n - 1))</td>
<td>(MS_W = \frac{SS_W}{JK(n - 1)})</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>(SS_T)</td>
<td>(N - 1)</td>
<td>()</td>
<td></td>
</tr>
</tbody>
</table>

### A Measure of Association (\(\omega^2\), Omega Squared)

Omega squared is a measure of association between the independent and dependent variables in an ANOVA statistic. The interpretation of \(\omega^2\) is analogous to the interpretation of \(r^2\) (coefficient of determination) – that is, \(\omega^2\) indicates the proportion of variance in the dependent variable that is accounted for by the levels of the independent variable.

The use of \(\omega^2\) can be extended to two-way ANOVA; the general formula is

\[
\omega^2 = \frac{SS_{\text{effect}} - (df_{\text{effect}})MS W}{SS_T + MS W}
\]

For the \((J)\) row main effect, the formula is:

\[
\omega^2 = \frac{SS_J - (J - 1)MS W}{SS_T + MS W}
\]

For the \((K)\) column main effect, the formula is:

\[
\omega^2 = \frac{SS_K - (K - 1)MS W}{SS_T + MS W}
\]

For the \((JK)\) interaction, the formula is:

\[
\omega^2 = \frac{SS_{JK} - (J - 1)(K - 1)MS W}{SS_T + MS W}
\]

Suppose for example that we obtained \(\omega^2 = .13\) for the row main effect (gender), \(\omega^2 = .44\) for the column main effect (program length), and \(\omega^2 = .13\) for the interaction... and our dependent variable was a flexibility measure. These data indicate that approximately 13 percent of the variance of the scores on the flexibility measure (dependent variable) can be attributed to the differences between males and females (row independent variable); approximately 44 percent can be attributed to the differences in the length of exercise programs (column independent variable); and approximately 13 percent can be attributed to the interaction of the two independent variables.
THE MEANING OF SIMPLE MAIN EFFECTS

Recall that a main effect is that effect of a factor ignoring (averaging across) the other factor. Simple main effects (a.k.a., simple effect) are conditional on the level of the other variable. That is, a simple effect is the effect of one factor (independent variable) at each level of the other factor.

The analysis of simple effects can be an important technique for analyzing data that contain significant interactions. In a very real sense, it allows us to “tease apart” significant interactions. Recall that a significant interaction tells us that the effect of one independent variable depends on the level of the other independent variable. That is, we seldom look at simple effects unless a significant interaction is present.

As a general rule, you should only look at those simple effects in which you are interested. We can look at the J row simple effect (e.g., A) at each of the levels of the K column (e.g., B) – and/or – the K column simple effect (B) at each of the levels of the J row (A). For example, if we have a 2 × 3 ANOVA, our simple effects could look like the following:

<table>
<thead>
<tr>
<th>Simple Effects of A</th>
<th>Simple Effects of B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A at B₁</td>
<td>B at A₁</td>
</tr>
<tr>
<td>A at B₂</td>
<td>B at A₂</td>
</tr>
<tr>
<td>A at B₃</td>
<td></td>
</tr>
</tbody>
</table>

To test the simple main effects, we must use the error term from the overall analysis (\(MS_w\), the estimate of error variance). \(MS_{error}\) continues to be based on all the data because it is a better estimate with more degrees of freedom. The sums of squares (SS) for the simple effects are calculated the same way as any sum of squares, using only the data that we are interested in. The degrees of freedom (df) for the simple effects are calculated in the same way as for the corresponding main effects sense the number of means we are comparing are the same. Post hoc follow-up procedures (dependent on the number of levels) are needed to understand the meaning of the simple effect results. We must continue to parcel out our calculations until we get to pairwise comparisons (i.e., \(X_i\) compared to \(X_k\)).

When we find significant pairwise differences – we will need to calculate an effect size for each of the significant pairs, which will need to be calculated by hand. An examination of the group means will tell us which group performed significantly higher than the other did.

For example, using the following formula:

\[ ES = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_w}} \]

Note that \(\bar{X}_i - \bar{X}_j\) (which can also be written as \(\bar{X}_i - \bar{X}_k\)) is the mean difference of the two groups (pair) under consideration. This value can be calculated by hand or found on the Contrast Results (K Matrix) table → looking at the value of the Difference (Estimate – Hypothesized). \(MS_w\) is the Within Group’s Mean Square value (a.k.a. Mean Square Within or ERROR), which is found on the ANOVA Summary Table (or Test Results Table).
TESTING THE NULL HYPOTHESIS FOR THE ONE-WAY ANOVA

1. State the hypotheses (null, $H_0$ and alternative, $H_a$)
   There is a null and alternative hypothesis for each main effect and the interaction

2. Set the criterion for rejecting $H_0$ (initial alpha level)

3. Check the assumptions
   Independence (design criteria)
   Normality (checked with Shapiro-Wilks, Standardized Skewness, etc.)
   Homogeneity of Variance (checked with Levene’s test of Homogeneity and Variance Ratio)

4. Compute the test statistic ($F$)
   Row ($J$) main effect
   Column ($K$) main effect
   Interaction ($JK$) effect

5. Decide whether to reject or retain $H_0$
   - If Interaction is significant $\rightarrow$ follow-up with Simple Main Effects
     - If Simple Main Effect is significant (controlling for Type I error) and has more than 2 levels $\rightarrow$ follow-up with Post hoc Procedures for pairwise comparisons
       - If Pairwise Comparison is significant (controlling for Type I error) $\rightarrow$ follow-up with Effect Size (using appropriate error term)
     - If Simple Main Effect is significant (controlling for Type I error) and has 2 levels $\rightarrow$ review group means $\rightarrow$ follow-up with Effect Size (using appropriate error term)
   - If Interaction is not significant $\rightarrow$ interpret main effect(s)
     - If Main Effect(s) is(are) significant and has more than 2 levels $\rightarrow$ follow-up with Post hoc Procedures for pairwise comparisons
       - If Pairwise Comparison is significant $\rightarrow$ follow-up with Effect Size (using appropriate error term)
     - If Main Effect(s) is(are) significant and has 2 levels $\rightarrow$ review group means $\rightarrow$ follow-up with Effect Size (using appropriate error term)

6. Calculate Measure of Association – as appropriate – and interpret

7. Interpret the results

8. Write a Results Section based on the findings
Results

A two-factor (2 x 3) Analysis of Variance was conducted to evaluate the effects of the length of an exercise program on the flexibility of female and male participants. The two independent variables in this study are gender and length of exercise program (1-week, 2-weeks, and 3-weeks). The dependent variable is the score on the flexibility measure, with higher scores indicating higher levels of flexibility. The means and standard deviations for the flexibility measure as a function of the two factors are presented in Table 1.

Table 1

*Means and Standard Deviations of Flexibility Levels*

<table>
<thead>
<tr>
<th></th>
<th>1-Week</th>
<th>2-Weeks</th>
<th>3-Weeks</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>24.63</td>
<td>27.37</td>
<td>41.00</td>
<td>31.00</td>
</tr>
<tr>
<td></td>
<td>(4.24)</td>
<td>(3.16)</td>
<td>(4.32)</td>
<td>(8.23)</td>
</tr>
<tr>
<td>Male</td>
<td>21.88</td>
<td>27.13</td>
<td>28.25</td>
<td>25.75</td>
</tr>
<tr>
<td></td>
<td>(4.76)</td>
<td>(2.90)</td>
<td>(3.28)</td>
<td>(4.56)</td>
</tr>
<tr>
<td>Total</td>
<td>23.25</td>
<td>27.25</td>
<td>34.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.58)</td>
<td>(2.93)</td>
<td>(7.56)</td>
<td></td>
</tr>
</tbody>
</table>

* Standard Deviations shown in parentheses

The test for normality, examining standardized skewness and the Shapiro-Wilks test indicated the data were statistically normal. The test for homogeneity of variance was not significant, Levene $F(5, 42) = .67, p = .649$, indicating that this assumption underlying the application of the two-way ANOVA was met. An alpha level of .05 was used for the initial analyses. The results for the two-way ANOVA indicated a significant main effect for gender, $F(1, 42) = 22.37, p < .001$ and a significant main effect for length of exercise program, $F(2, 42)$
= 36.03, \( p < .001 \). Additionally, the results show a significant interaction between gender and length of exercise program, \( F(2, 42) = 11.84, p < .001 \) (see Table 2), indicating that any differences between the length of exercise programs were dependent upon which gender the subjects were and that any differences between females and males were dependent upon which length of exercise program they were in (see Figure 1 for a graph of this interaction).

Approximately 14\% (\( \omega^2 = .135 \)) of the total variance of the flexibility levels was attributed to the interaction of gender and length of exercise program.

Table 2

Two-way Analysis of Variance for Flexibility Levels

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>330.75</td>
<td>1</td>
<td>330.75</td>
<td>22.37</td>
<td>.000</td>
</tr>
<tr>
<td>Length of Program</td>
<td>1065.50</td>
<td>2</td>
<td>532.75</td>
<td>36.03</td>
<td>.000</td>
</tr>
<tr>
<td>Gender ( \times ) Length</td>
<td>350.00</td>
<td>2</td>
<td>175.00</td>
<td>11.84</td>
<td>.000</td>
</tr>
<tr>
<td>Within (Error)</td>
<td>621.00</td>
<td>42</td>
<td>14.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2367.25</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TWO-WAY ANOVA
PAGE 12
Because the interaction between gender and the length of exercise program was significant, we chose to ignore the two main effects and instead first examined the gender simple main effects, that is, the differences between females and males for each of the three lengths of exercise programs. To control for Type I error rate across the three simple effects, we set the alpha level for each at .0167 ($\alpha/3 = .05/3$). The only significant difference between females and males was found in the 3-weeks exercise program. A review of the group means indicated that females ($M = 41.00$) had a significantly higher level of flexibility than males ($M = 28.25$), $F(1, 42) = 43.98, p < .001, ES = 3.32$.

Additionally, we examined the length of exercise program simple main effects, that is, the differences among the three lengths of exercise programs for females and males separately. To control for Type I error across the two simple main effects, we set the alpha level for each at
There was a significant difference among the three lengths of exercise programs for females, $F(2, 42) = 41.60, p < .001$, and for males, $F(2, 42) = 6.26, p < .01$. Follow-up tests were conducted to evaluate the three lengths of exercise programs’ pairwise differences for females. The alpha level was set at $.0083 (.025/3)$ to control for Type I error over the three pairwise comparisons. The females in the 3-weeks exercise program ($M = 41.00$) had significantly higher flexibility levels compared to the females in the 1-week program ($M = 24.63$), $F(1, 42) = 72.54, p < .001$, $ES = 4.26$ and the females in the 2-weeks program ($M = 27.37$), $F(1, 42) = 50.22, p < .001$, $ES = 3.54$. Follow-up tests were also conducted to evaluate the three lengths of exercise programs’ pairwise differences for males. The alpha level was set at $.0083 (.025/3)$ to control for Type I error over these three pairwise comparisons. The males in the 3-weeks exercise program ($M = 28.25$) had significantly higher flexibility levels compared to the males in the 1-week program ($M = 21.88$), $F(1, 42) = 11.00, p < .008$, $ES = 1.66$.

REFERENCES


