UNDERSTANDING THE ONE-WAY ANOVA

The One-way Analysis of Variance (ANOVA) is a procedure for testing the hypothesis that K population means are equal, where $K \geq 2$. The One-way ANOVA compares the means of the samples or groups in order to make inferences about the population means. The One-way ANOVA is also called a *single factor analysis of variance* because there is only one independent variable or factor. The independent variable has nominal levels or a few ordered levels.

PROBLEMS WITH USING MULTIPLE t TESTS

To test whether pairs of sample means differ by more than would be expected due to chance, we might initially conduct a series of t tests on the K sample means – however, this approach has a major problem (i.e., inflating the Type I Error rate).

The number of t tests needed to compare all possible pairs of means would be $K(K-1)/2$, where K = number of means.

When more than one t test is run, each at a specified level of significance (such as α = .05), the probability of making one or more Type I errors in the series of t tests is greater than α.

The **Type I Error Rate** is determined as $1-(1-\alpha)^c$

Where α = level of significance for each separate t test

c = number of independent t tests

Sir Ronald A. Fisher developed the procedure known as analysis of variance (ANOVA), which allows us to test the hypothesis of equality of K population means while maintaining the Type I error rate at the pre-established (*a priori*) α level for the entire set of comparisons.

THE VARIABLES IN THE ONE-WAY ANOVA

In an ANOVA, there are two kinds of variables: **independent** and **dependent**.

The **independent variable** is controlled or manipulated by the researcher. It is a **categorical** (*discrete*) variable used to form the groupings of observations. However, do not confuse the independent variable with the "levels of an independent variable."

In the One-way ANOVA, **only one independent variable is considered**, but there are two or more (theoretically any finite number) levels of the independent variable. The independent variable is typically a categorical variable. The independent variable (or factor) divides individuals into two or more groups or levels. The procedure is a One-way ANOVA, since there is only one independent variable.

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.

The (**continuous**) *dependent variable* is defined as the variable that is, or is presumed to be, the result of manipulating the independent variable.

In the One-way ANOVA, there is only one dependent variable – and hypotheses are formulated about the means of the groups on that dependent variable. The dependent variable differentiates individuals on some quantitative (continuous) dimension.

The ANOVA *F* test (named after Sir Ronald A. Fisher) evaluates whether the group means on the dependent variable differ significantly from each other. That is, an overall analysis-of-variance test is conducted to assess whether means on a dependent variable are significantly different among the groups.

MODELS IN THE ONE-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. For a one-way analysis of variance, the distinction is not particularly critical, but it can become quite important when working with more complex designs such as the factorial analysis of variance.

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 in single-factor designs. 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.

HYPOTHESES FOR THE ONE-WAY ANOVA

The **null hypothesis** (H_0) tested in the One-way ANOVA is that the population means from which the K samples are selected are equal. Or that each of the group means is equal.

H_0: $\mu_1 = \mu_2 = ... = \mu_K$ Where K is the number of levels of the independent variable.

For example: If the independent variable has three levels – we would write…

H_0: $\mu_1 = \mu_2 = \mu_3$

If the independent variable has five levels – we would write…

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

The subscripts could be replaced with group indicators.

For example: H_0: $\mu_{Method1} = \mu_{Method2} = \mu_{Method3}$

The **alternative hypothesis** (H_a) is that at least one group mean significantly differs from the other group means. Or – that at least two of the group means are significantly different.

H_a: $\mu_i \neq \mu_k$ for some i, k Where i and k simply indicate unique groups.

ASSUMPTIONS UNDERLYING THE ONE-WAY ANOVA

1. The observations are random and independent samples from the populations. This is commonly referred to as the **assumption of independence**.

    The null hypothesis actually says that the samples come from populations that have the same mean. The samples must be random and independent if they are to be representative of the populations. The value of one observation is not related to any other observation. In other words, one person's score should not provide any clue as to how any of the other people should score. That is, one event does not depend on another.

2. The distributions of the populations from which the samples are selected are normal. This is commonly referred to as the **assumption of normality**.

    This assumption implies that the dependent variable is normally distributed (a theoretical requirement of the underlying distribution, the F distribution) in each of the groups.

3. The variances of the distributions in the populations are equal. This is commonly referred to as the **assumption of homogeneity of variance**.

    This assumption (along with the normality assumption and the null hypothesis) provides that the distributions in the populations have the same shapes, means, and variances; that is, they are the same populations. In other words, the variances on the dependent variable are equal across the groups.

    Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_K^2$ (if retained = assumption met)

                                                        (if rejected = assumption not met)

    Alternative Hypothesis: H_a: $\sigma_i^2 \neq \sigma_k^2$ for some i, k

TESTING THE ASSUMPTION OF INDEPENDENCE

One of the first steps in using the One-way analysis of variance is to test the assumption of independence. Random sampling and independence are technical concerns related to the way the samples are selected. Independence is a methodological concern; which is dealt with (or should be dealt with) when a study is set up. Although the independence assumption can ruin a study if it is violated, there is no way to use the study's sample data to test the validity of this prerequisite condition. It is assessed through an examination of the design of the study. That is, we confirm that the K groups are independent of each other?

The assumption of independence is commonly known as the unforgiving assumption (r.e., robustness), which simply means that if the *K* groups are not independent of each other, one cannot use the one-way analysis of variance. That is, if the groups (categories) are independent of each other, which is typically assessed through an examination of the research design, the assumption of independence has been met. If this assumption is not met, the one-way ANOVA is an inappropriate statistic.

TESTING THE ASSUMPTION OF NORMALITY

Another of the first steps in using the One-way ANOVA test is to test the assumption of normality, where the Null Hypothesis is that there is no significant departure from normality for each of the groups/levels, as such; retaining the null hypothesis indicates that the assumption of normality has been met for the given samples. The Alternative Hypothesis is that there is a significant departure from normality, as such; rejecting the null hypothesis in favor of the alternative indicates that the assumption of normality has not been met for the given samples.

The normality assumption can also be checked through an examination of the research design. It can be tested through supplemental statistical analyses, such as a "goodness-of-fit" test (e.g., Chi-Square), which tests whether the sample distributions are within random sampling fluctuation of normality.

To test the assumption of normality, we can use the Shapiro-Wilks test, which is more commonly used by statisticians. The Shapiro-Wilks Test is a statistical test of the hypothesis that sample data have been drawn from a normally distributed population. From this test, the Sig. (*p*) value is compared to the *a priori* alpha level (level of significance for the statistic) – and a determination is made as to reject ($p \leq \alpha$) or retain ($p > \alpha$) the null hypothesis.

Tests of Normality

	style Attachment Style	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
delta Delta Sleep	1 Secure	.182	10	.200*	.930	10	.445
	2 Anxious	.253	10	.068	.915	10	.314
	3 Avoidant	.164	10	.200*	.968	10	.876

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

For the above example, where α = .001, given that *p* = .445 for the Secure Group, *p* = .314 for the Anxious Group, and *p* = .876 for the Avoidant Group – we would conclude that each of the levels of the Independent Variable (Attachment Style) are normally distributed. Therefore, the assumption of normality has been met for this sample. The *a priori* alpha level is based on sample size – where .05 and .01 are commonly used. Tabachnick and Fidell (2007) report that conventional but conservative (.01 and .001) alpha levels are commonly used to evaluate the assumption of normality.

NOTE: Since the Shapiro-Wilks test is rather conservative, most statisticians will agree that the Shapiro-Wilks Test should not be the sole determination of normality. We typically supplement our assessment of normality with an examination of skewness (in excess of +3.29 is a concern), kurtosis (in excess of +3.29 is a concern), an examination of the

histogram graph (non-normal shape is a concern), an examination of box plots (extreme outliers is a concern), and an examination of the Normal Q-Q Plots (a non-linear relationship is a concern). These procedures are typically done during data screening.

In examining skewness and kurtosis, we divide the skewness (kurtosis) statistic by its standard error. We want to know if this standard score value significantly departs from normality. Concern arises when the skewness (kurtosis) statistic divided by its standard error is greater than $z \pm 3.29$ ($p < .001$, two-tailed test) (Tabachnick & Fidell, 2007).

We have several options for handling non-normal data, such as deletion and data transformation (based on the type and degree of violation as well as the randomness of the missing data points). Any adjustment to the data should be justified (i.e., referenced) based on solid resources (e.g., prior research or statistical references). As a first step, data should be thoroughly screened to ensure that any issues are not a factor of missing data or data entry errors. Such errors should be resolved prior to any data analyses using acceptable procedures (see for example Howell, 2007 or Tabachnick & Fidell, 2007).

TESTING THE ASSUMPTION OF HOMOGENEITY-OF-VARIANCE

Another one of the first steps in the One-way ANOVA test is to test the assumption of homogeneity of variance, where the null hypothesis assumes no difference between the K group's variances. One method is the Bartlett's test for homogeneity of variance (this test is very sensitive to non-normality).

The Levene's F Test for Equality of Variances, which is the most commonly used statistic (and is provided in SPSS), is used to test the assumption of homogeneity of variance. Levene's test uses the level of significance set *a priori* for the ANOVA (e.g., $\alpha = .05$) to test the assumption of homogeneity of variance.

Test of Homogeneity of Variances

SCORE

Levene Statistic	df1	df2	Sig.
1.457	2	42	.244

For Example: For the SCORE variable (shown above), the F value for Levene's test is 1.457 with a Sig. (p) value of .244. Because the Sig. value is greater than our alpha of .05 ($p > .05$), we retain the null hypothesis (no difference) for the assumption of homogeneity of variance and conclude that there is not a significant difference between the three group's variances. That is, the assumption of homogeneity of variance is met.

Test of Homogeneity of Variances

VISUAL

Levene Statistic	df1	df2	Sig.
17.570	1	498	.000

For Example: For the VISUAL variable (shown above), the *F* value for Levene's test is 17.570 with a Sig. (*p*) value of .000 (< .001). Because the Sig. value is less than our alpha of .05 ($p < .05$), we reject the null hypothesis (no difference) for the assumption of homogeneity of variance and conclude that there is a significant difference between the two group's variances. That is, the assumption of homogeneity of variance is **not** met.

VIOLATION OF THE ASSUMPTIONS OF THE ONE-WAY ANALYSIS OF VARIANCE

If a statistical procedure is little affected by violating an assumption, the procedure is said to be **robust** with respect to that assumption. The One-way ANOVA is robust with respect to violations of the assumptions, except in the case of unequal variances with unequal sample sizes. That is, the ANOVA can be used when variances are only approximately equal if the number of subjects in each group is equal (where equal can be defined as the larger group size not being more then 1½ times the size of the smaller group). ANOVA is also robust if the dependent variable data are even approximately normally distributed. Thus, if the assumption of homogeneity of variance (where the larger group variance is not more than 4 or 5 times that of the smaller group variance), or even more so, the assumption of normality is not fully met, you may still use the One-way ANOVA.

Generally, failure to meet these assumptions changes the Type I error rate. Instead of operating at the designated level of significance, the actual Type I error rate may be greater or less than α, depending on which assumptions are violated.

When the population sampled are not normal, the effect on the Type I error rate is minimal.

If the population variances differ, there may be a serious problem when sample sizes are unequal.

If the larger variance is associated with the larger sample, the *F* test will be too conservative.

If the smaller variance is associated with the larger sample, the *F* test will be too liberal. (If the α level is .05, "conservative" means that the actual rate is less than .05.)

If the sample sizes are equal, the effect of heterogeneity of variances (i.e., violating the assumption of homogeneity of variance) on the Type I error is minimal. In other words, the effects of violating the assumptions vary somewhat with the specific assumptions violated.

If there are extreme violations of these assumptions – with respect to normality and homogeneity of variance – an alternate test such as the *Kruskal-Wallis* test should be used instead of the one-way analysis of variance test.

The Kruskal-Wallis test is a nonparametric test that is used with an independent groups design employing *K* groups. It is used as a substitute for the parametric one-way ANOVA, when the assumptions of that test are seriously violated. The Kruskal-Wallis test does not assume population normality nor homogeneity of variance, as does the parametric ANOVA, and only requires ordinal scaling of the dependent variable. It is used when violations of population normality and/or homogeneity of variance are extreme or when interval or ratio scaling are required and not met by the data.

THE WELCH AND BROWN-FORSYTHE STATISTIC

If the equal variance assumption has been violated (e.g., if the significance for the Levene's test is less than .05), we can use an adjusted F statistic. Two such types of adjustments are provided by the Welch statistic and the Brown-Forsythe statistic. The Welch test is more powerful and more conservative than the Brown-Forsythe test.

If the F ratio is found to be significant with either the Welch statistic or the Brown-Forsythe statistic, an appropriate post hoc test would be required. The Games-Howell post hoc test, for example, is appropriate when the equal variances assumption has been violated. This test is not appropriate if equal variance were assumed.

Robust Tests of Equality of Means

Number of words recalled

	Statistic[a]	df1	df2	Sig.
Welch	9.037	4	21.814	.000
Brown-Forsythe	6.095	4	18.706	.003

a. Asymptotically F distributed.

The output from the above table is only valid if the equal variance assumption has been violated. From this example, using the Welch statistic, we find that $F(4, 21.814) = 9.037$, $p < .001$. If for example, our *a priori* alpha level were set at .05, we would conclude that the adjusted F ratio is significant. Since the p value is smaller than α we would reject the null hypothesis and we would have permission to proceed and compare the group means.

The difference between the adjusted F ratio (devised by Welch and Brown and Forsythe) and the ordinary F ratio is quite similar to that of the adjusted t and ordinary t found in the independent-samples t test. In both cases it is only the denominator (i.e., error term) of the formula that changes.

SUMMARY TABLE FOR THE ONE-WAY ANOVA

Summary ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance Estimate (Mean Square)	F Ratio
Between	SS_B	$K - 1$	$MS_B = \dfrac{SS_B}{K-1}$	$\dfrac{MS_B}{MS_W}$
Within	SS_W	$N - K$	$MS_W = \dfrac{SS_W}{N-K}$	
Total	$SS_T = SS_B + SS_W$	$N - 1$		

The analysis of variance partitions the total variability of the data (SS_T) into two sources: the variability that exists within each group, called the *within-groups sum of squares* (SS_W), and the variability that exists between the groups, called the *between-groups sum of squares* (SS_B). Each sum of squares is used to form an independent estimate of the H_0 population variance. The estimate based on the within-groups variability is called the *within-groups variance estimate* (also known as Mean Square Within, MS_W), and the estimate based on the between-group variability is called the *between-groups variance estimate* (also known as Mean Square Between, MS_B). Finally, an *F* ratio (F_{obt}) is calculated where

$$F_{obt} = \frac{MS_B}{MS_W}$$

This first component (MS_W) reflects the differences observed among subjects exposed to the same treatment. <u>It is assumed</u> that within-groups variation of a similar magnitude exists in each of the groups. This variation within any one group is a function of the specific subjects selected at random for the group or allocated at random to the group. Therefore, we can attribute variation within a group to random sampling fluctuation (i.e., sampling error – which is why MS_W is also referred to as ERROR).

The second component (MS_B) has to do with the differences among group means. Even if there were absolutely no treatment effects, it would be unlikely that the sample means for the groups would be identical. A more reasonable expectation is that the group means will differ, even without treatment, simply as a function of the individual difference of the subjects. Thus, <u>we would expect</u> the group means to vary somewhat due to the random selection (assignment) process in the formation of the groups. If, in addition, different treatments that do have an effect on the dependent variable are applied to the different groups, we can expect even larger differences among the group means. Thus, the between-groups variation reflects variation due to the treatment <u>plus</u> variation attributable to the random process by which subjects are selected and assigned to groups. That is, the treatment effect of the independent variable plus error.

Which means $$F_{obt} = \frac{\text{Treatment Effect} + \text{Error}}{\text{Error}}$$

When the null hypothesis is true (no difference between the group means – or – no treatment effect), we would expect *F* to be equal to 1. Note that the observed mean squares merely estimate the parameters and that these estimates may be larger or smaller than the corresponding parameters. Therefore, it is possible to have an observed *F* ratio less than 1.00, even though conceptually the ratio cannot be less than 1.00. *F* increases with the effect of the independent variable. Thus, the larger the *F* ratio is, the more reasonable it is that the independent variable has had a real effect. If the *F* ratio is less than 1.00, we don't even need to compare it with the critical value of *F* (F_{crit}). It is obvious that the treatment has not had a significant effect, and we can immediately conclude by retaining H_0.

A MEASURE OF ASSOCIATION (ω^2, OMEGA SQUARED)

By rejecting the null hypothesis in ANOVA – this indicates that there are significant differences somewhere among the sample means, a greater difference than would be expected on the basis of chance alone. However, with large sample sizes, these statistically significant differences may have little practical significance.

A measure of the strength of the association between the independent variable and the dependent variable in ANOVA is ω^2, omega squared. Omega squared indicates the proportion of the total variance in the dependent variable that is accounted for by the levels of the independent variable. This is analogous to the coefficients of determination (r^2).

The formula for omega squared is:

$$\omega^2 = \frac{SS_B - (K-1)MS_W}{SS_T + MS_W}$$

For example, if we calculated ω^2 = .3928, this means that the independent variable in the ANOVA accounts for approximately 39.28% of the total variance in the dependent variable.

POST HOC (MULTIPLE COMPARISON) PROCEDURES

Post hoc comparisons (also called post hoc tests, multiple comparison tests, a posteriori comparisons, and follow-up test or comparisons) are tests of the statistical significance of differences between group means calculated after ("post") having done an analysis of variance (ANOVA) that shows an overall difference. The *F* ratio of the ANOVA tells us that some sort of statistically significant differences exist somewhere among the groups being studied. Subsequent, post hoc, analyses are meant to specify what kind and where.

Post hoc comparison is a generic term for several kinds of analyses. Tukey's Honestly Significant Difference (HSD) test and the Scheffé method are the two best known. Others include the Newman-Keuls test and Duncan's Multiple-Range test. Of the four, Duncan's is the least conservative measure, and Scheffé's is the most conservative. Tukey's is more often associated with the ANOVA. Tukey's is designed to make all pairwise comparisons while maintaining the experimentwise error rate (α_E) at the pre-established α level.

There are numerous post hoc (multiple comparisons) procedure tests available. Your decision should be based on prior experience, the research situation, and/or the literature. If the assumption of homogeneity of variance has been met (Equal Variances Assumed) – the most commonly used test is the Tukey (HSD) test. If the assumption of homogeneity of variance has been violated (Equal Variances Not Assumed) – the Games-Howell or the Dunnett's C test are commonly used. Also keep in mind that if the equal variance assumption has not been met, we typically would adjust the *F* ratio using either the Welch statistic or the Brown-Forsythe statistic.

Note – when setting up the steps in our analysis – it is common to request either the Welch or Brown-Forsythe statistic and select at least one post hoc test for either situation… as a just-in-case (i.e., at the onset of our analysis – we do not know if the equal variance assumption has been met or violated).

Effect size, broadly, is any of several measures of association or of the strength of a relation, such as Pearson's *r* or eta (η). Effect size is thought of as a measure of practical significance. It is defined as the degree to which a phenomenon exists. Keep in mind that there are several acceptable measures of effect size. The choice should be made based on solid references based on the specific analysis being conducted.

So why bother with effect size at all? Any observed difference between two sample means can be found to be statistically significant when the sample sizes are sufficiently large. In such a case, a small difference with little practical importance can be statistically significant. On the other hand, a large difference with apparent practical importance can be nonsignificant when the sample sizes are small. Effect sizes provide another measure of the magnitude of the difference expressed in standard deviation units in the original measurement. Thus, with the test of statistical significance (e.g., the *F* statistic) and the interpretation of the effect size (*ES*), the researcher can address issues of both statistical significance and practical importance.

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:
$$ES = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{MS_W}}$$

Note that $\overline{X}_i - \overline{X}_j$ (which can also be written as $\overline{X}_i - \overline{X}_k$) is the mean difference of the two groups (pairs) under consideration. This value can be calculated by hand or found in the **Mean Difference (I-J)** column on the **Multiple Comparison** table in SPSS. 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.

Suppose that the mean for the Red Group = 16.60 and the mean for the Green Group = 11.10, and the Mean Square Within (MS_W) found in the ANOVA table = 16.136, we would find that the *ES* = 1.37. That is, the mean difference of 5.50 is 1.37 standard deviation units away from the hypothesized mean difference of 0. Recall that H_0: $\mu_1 - \mu_2 = 0$.

For Red / Green, we find $ES = \dfrac{16.60 - 11.10}{\sqrt{16.136}} = \dfrac{5.50}{4.016964} = 1.3691933 = \mathbf{1.37}$

TESTING THE NULL HYPOTHESIS FOR THE ONE-WAY ANOVA

In testing the null hypothesis for the One-way ANOVA – we follow these steps:

1. STATE THE HYPOTHESES

   - State the Null Hypothesis and Alternative Hypothesis

2. SET THE CRITERION FOR REJECTING H_0

   - Set the alpha level, which in turn identifies the critical values

3. TEST THE ASSUMPTIONS FOR THE ONE-WAY ANOVA

   - Assumption of Independence (e.g., Research Design)

   - Assumption of Normality (e.g., Shaprio-Wilks)

   - Assumption of Homogeneity of Variance (e.g., Levene's Test of Homogeneity)

4. COMPUTE THE TEST STATISTIC

   - Calculate the F ratio (or adjusted F ratio using for example the Welch statistic)

5. DECIDE WHETHER TO RETAIN OR REJECT H_0

   - Using the obtained probability value (compared to the alpha level)

     - If $p > \alpha$ → Retain the null hypothesis

     - If $p \leq \alpha$ → Reject the null hypothesis

   - Using the obtained statistic value (compared to the critical value)

     - If $F_{obt} < F_{crit}$ → Retain the null hypothesis

     - If $F_{obt} \geq F_{crit}$ → Reject the null hypothesis

6. CALCULATE MEASURE OF ASSOCIATION (ω^2, OMEGA SQUARED)

   - If the null hypothesis is rejected, calculate omega squared to determine strength of the association between the independent variable and the dependent variable.

7. DECIDE WHETHER TO RETAIN OR REJECT H_0 FOR EACH OF THE PAIRWISE COMPARISONS (I.E., CONDUCT POST HOC PROCEDURES)

   - If the null hypothesis is rejected, use the appropriate post hoc procedure to determine whether unique pairwise comparisons are significant.

   - Choice of post hoc procedures is based on whether the assumption of homogeneity of variance was met (e.g., Tukey HSD) or not (e.g, Games-Howell).

8. CALCULATE EFFECT SIZE

   - Calculate an effect size for each significant pairwise comparison

9. INTERPRET THE RESULTS

10. WRITE A RESULTS SECTION based on the findings

A One-way Analysis of Variance (ANOVA) was used to examine whether students' scores on a standardized test is a function of the teaching method they received. The independent variable represented the three different types of teaching methods: 1) lecture only; 2) hands-on only; and 3) lecture and hands-on. The dependent variable was the students' score on a standardized test. See Table 1 for the means and standard deviations for each of the three groups.

Table 1

Means and Standard Deviations of Standardized Test Scores

Method	*n*	*Mean*	*SD*
Lecture Only	15	82.80	9.59
Hands-On Only	15	88.53	8.73
Lecture and Hands-On	15	92.67	6.22
Total Group	45	88.00	9.09

An alpha level of .05 was used for all analyses. The test for homogeneity of variance was not significant [*Levene F*(2, 42) = 1.46, *p* > .05] indicating that this assumption underlying the application of ANOVA was met. The one-way ANOVA of standardized test score (see Table 2) revealed a statistically significant main effect [*F*(2, 42) = 5.34, *p* < .01] indicating that not all three groups of the teaching methods resulted in the same standardized test score. The ω^2 = .162 indicated that approximately 16% of the variation in standardized test score is attributable to differences between the three groups of teaching methods.

Table 2

Analysis of Variance for Standardized Test Scores

Source	*SS*	*df*	*MS*	*F*	*p*
Between	736.53	2	368.27	5.34	.009
Within	2895.47	42	68.94		
Total	3632.00	44			

Post hoc comparisons using Tukey procedures were used to determine which pairs of the three group means differed. These results are given in Table 3 and indicate that students who had received the lecture and hands-on teaching method (M = 92.67) scored significantly higher on the standardized test than did students who had received the lecture only teaching method (M = 82.80). The effect size for this significant pairwise difference was 1.19.

Table 3

Tukey Post Hoc Results and Effect Size of Standardized Test Scores by Teaching Method

Group	Mean	Mean Differences ($\overline{X}_i - \overline{X}_k$) (Effect Size is indicated in parentheses)		
		1.	2.	3.
1. Lecture Only	82.80	0.00		
2. Hands-On Only	88.53	5.73	0.00	
3. Lecture and Hands-On	92.67	9.87** (1.19)	4.13	0.00

** $p < .01$

REFERENCES

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). New York: Houghton Mifflin Company.

Howell, D. C. (2002). *Statistical Methods for Psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2004). *SPSS for Introductory Statistics: Use and Interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Pagano, R. R. (2004). *Understanding Statistics in the Behavioral Sciences* (7th ed.). Belmont, CA: Thomson/Wadsworth.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.