

## INTERPRETING THE ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

As with other parametric statistics, we begin the one-way ANOVA with a test of the underlying assumptions. Our first assumption is the assumption of independence. Recall that this assumption is assessed through an examination of the design of the study. That is, we confirm that the  $K$  groups/levels are independent of each other.

We must also test the assumption of normality for the  $K$  levels of the independent variable. To test the assumption of normality, we can use the Shapiro-Wilks test, which is commonly used by statisticians, and is typically tested at the  $\alpha = .001$  level of significance. 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. As a general rule, we should use other measures of normality checking in conjunction with the Shapiro-Wilks test (e.g., standardized skewness).

For our example, we obtained the following results:

Tests of Normality							
Treatment Group		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Gain Score	Red Group	.149	10	.200*	.946	10	.627
	Blue Group	.184	10	.200*	.918	10	.341
	Green Group	.139	10	.200*	.948	10	.643
	Yellow Group	.125	10	.200*	.958	10	.766

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

a. Lilliefors Significance Correction

Given that  $p = .627$  for the Red Group,  $p = .341$  for the Blue Group,  $p = .643$  for the Green Group, and  $p = .766$  for the Yellow Group – and using  $\alpha = .001$  – we would conclude that each of the levels of the Independent Variable (Treatment Group) are normally distributed. Therefore, the assumption of normality has been met for this sample.

The first table from the ANOVA output, (**DESCRIPTIVES**) provides familiar descriptive statistics (e.g., Group Size, Mean, Standard Deviation) for the four color groups on the dependent variable that we requested (Gain Score) for our example.

The second table from the ANOVA output, (**TEST OF HOMOGENEITY OF VARIANCES**) provides the Levene's Test to check the assumption that the variances of the four color groups are equal; i.e., not significantly different. Notice that the Levene's test is not significant;  $F(3, 36) = 1.485$ ,  $p = .235$  – at the .05 alpha level for our example. Thus, the assumption of homogeneity of variance is met (i.e., not violated) for this sample.

If the result had been significant (i.e., assumption not met), then we could use an adjusted  $F$  test such as the Welch statistic or the Brown-Forsythe statistic. If there are extreme violations to the assumption of normality and the assumption of homogeneity of variance, a nonparametric test such as Kruskal-Wallis could be used.

The third table from the ANOVA output, (**ANOVA**) is the key table because it shows whether the overall  $F$  ratio for the ANOVA is significant. Note that our  $F$  ratio (6.414) is significant ( $p = .001$ ) at the .05 alpha level. When reporting this finding – we would write, for example,  $F(3, 36) = 6.41, p < .01$ . The **F** indicates that we are using an  $F$  test (i.e., ANOVA). The **3** and **36** are the two degrees of freedom values ( $df$ ) for the between groups “effect” and the within-groups “error,” respectively. The **6.41** is the obtained  $F$  ratio, and the  **$p < .01$**  is the probability of obtaining that  $F$  ratio by chance alone.  $F$  tables also usually include the mean squares, which indicates the amount of variance (sums of squares) for that “effect” divided by the degrees of freedom for that “effect.” We also should report the means and standard deviations so that our readers can see what groups were low and high. Remember, however, that if you have three or more groups (such as in our example) we will not know which specific pairs of means are significantly different, unless we do a post hoc test or an *a priori* comparison test.

**At this point** – we have rejected the null hypothesis that all four groups’ means are equal, since  $p \leq \alpha$ . We conclude that at least one of the group means is significantly different from the others (or that at least two of the group means are significantly different from each other). Beyond this conclusion – we will need to conduct a post hoc follow-up test to determine which means differ from each other.

Since we have a significant  $F$ , we will want to measure the strength of association ( $\omega^2$ ) between the independent variable and the dependent variable for our example. Note that the omega square ( $\omega^2$ ) is not a paired effect size. This will need to be calculated by hand – SPSS does not provide omega square values.

Using the following formula:

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

$$\text{We find } \omega^2 = \frac{310.475 - (4 - 1)16.136}{891.375 + 16.136} = \frac{262.067}{907.511} = 0.2887755 = \mathbf{.29}$$

Which indicates that the independent variable (four color groups) accounts for approximately 29% of the variance in the dependent variable (gain score) for this sample.

The fourth table from the ANOVA output, (**ROBUST TESTS OF EQUALITY OF MEANS**) is considered when the assumption of homogeneity of variance has not been met. The adjusted  $F$  ratio and its applicable Sig. ( $p$ ) value are provided. If the adjusted  $F$  ratio is found to be significant (i.e.,  $p \leq \alpha$ ), we would reject the null hypothesis and conclude that at least one of the group means is significantly different from the others (or that at least two of the group means are significantly different from each other). Beyond this conclusion – we will need to conduct a post hoc follow-up test. The applicable post hoc analysis will have to take into account that the equal variance assumption has been violated (e.g., Games-Howell). Note – when setting up the steps in our analysis – it is common to select at least one adjusted  $F$  ratio statistics... as a just-in-case (i.e., at the onset of our analysis – we do not know if the assumption has been met or violated). For this example, since the equal variance assumption has been met, we can ignore the information in this table.

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. Note – when setting up the steps in our analysis – it is common to 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 assumption has been met or violated).

## INTERPRETING THE POST HOC TESTS

The **MULTIPLE COMPARISONS** table (in our example) is showing the results for the **Tukey HSD** and the **Games-Howell** follow-up tests. Since the assumption of homogeneity of variance was met in our example – we only need to review the Tukey HSD information. The information for the Games-Howell can be ignored at this time. At first glance, this table is rather intimidating – however, there is only certain pieces of data that we need to make our conclusion. First locate the **(I) GROUP** (let's start with **1 Red Group**) – then locate the **(J) GROUP** (let's look at the first line, **2 Blue Group**) – this is the Red Group being compared to the Blue Group. In the next column, **Mean Difference (I-J)** – we find the mean difference for these two groups (i.e., Red Group's mean = 16.60 and Blue Group's mean = 12.80, therefore  $16.60 - 12.80 = 3.80$ ). The next column of interest is the **Sig.** column, which tells us the  $p$  value ( $p = .168$  in this case). The  $p$  value is then compared to the alpha level to make the determination of whether this pair is significantly different. Notice that there is a degree of redundancy in this table – so we are only interested in the unique pair comparisons. Further review of the table reveals that the Red Group ( $M = 16.60$ ) is significantly different from the Green Group ( $M = 11.10$ ), with a mean difference of 5.50 and a  $p$  value of .021. Also, the Red Group ( $M = 16.60$ ) is significantly different from the Yellow Group ( $M = 9.00$ ), with a mean difference of 7.60 and a  $p$  value of .001.

Since we found significant pairwise differences – we will need to calculate an effect size for each of these 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.

Using the following formula:

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

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

$$\text{For Red / Yellow, we find } ES = \frac{16.60 - 9.00}{\sqrt{16.136}} = \frac{7.60}{4.016964} = 1.8919761 = \mathbf{1.89}$$

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 (pairs) under consideration. This value can be calculated by hand or found in the **Mean Difference (I-J)** column on the **Multiple Comparison** table.  $MS_W$  is the Within Group's Mean Square value (a.k.a. Mean Square Within or ERROR), which is found on the ANOVA Summary Table.

The **HOMOGENEOUS SUBSETS** table is also provided by SPSS (shown below). This table provides an alternative way of computing and displaying the post hoc tests, and is considered more appropriate when group sizes are quite different. Groups listed in the same subset are not significantly different. So, for our example, the Gain Scores of the Yellow Group, the Green Group, and the Blue Group are not significantly different. Likewise, the Gain Scores of the Blue Group and the Red Group are not significantly different.

## Homogeneous Subsets

Gain Score				
Treatment Group		N	Subset for alpha = .05	
			1	2
Tukey HSD <sup>a</sup>	Yellow Group	10	9.00	
	Green Group	10	11.10	
	Blue Group	10	12.80	12.80
	Red Group	10		16.60
	Sig.		.168	.168

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

We can also request a **MEANS PLOT** from SPSS, which will give us a visual representation of the group means and their linear relationship. Such a line graph can aid in interpreting the results. The plot may need to be adjusted (re-scaled) to provide a clearer picture of the linear relationship between the variables. That is, the plot may appear to show notable difference between the groups, but when put on the applicable scale (from the dependent variable), the difference may be reduced... however, more appropriate based on the applicable scale.

# INTERPRETING THE ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

## Oneway ANOVA (In-Class Example)

### Descriptives

GAIN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 Red Group	10	16.60	4.452	1.408	13.42	19.78	10	23
2 Blue Group	10	12.80	5.116	1.618	9.14	16.46	7	23
3 Green Group	10	11.10	2.961	.936	8.98	13.22	5	15
4 Yellow Group	10	9.00	3.127	.989	6.76	11.24	5	15
Total	40	12.38	4.781	.756	10.85	13.90	5	23

### Test of Homogeneity of Variances

GAIN

Levene Statistic	df1	df2	Sig.
1.485	3	36	<b>.235</b>

### ANOVA

GAIN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	310.475	3	103.492	6.414	<b>.001</b>
Within Groups	580.900	36	16.136		
Total	891.375	39			

### Robust Tests of Equality of Means

Gain Score

	Statistic <sup>a</sup>	df1	df2	Sig.
Welch	6.348	3	19.605	.003
Brown-Forsythe	6.414	3	29.979	.002

a. Asymptotically F distributed.

## Post Hoc Tests

### Multiple Comparisons

Dependent Variable: GAIN

			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
(I) GROUP	(J) GROUP	Lower Bound				Upper Bound	
Tukey HSD	1 Red Group	2 Blue Group	3.80	1.796	.168	-1.04	8.64
		3 Green Group	5.50*	1.796	.021	.66	10.34
		4 Yellow Group	7.60*	1.796	.001	2.76	12.44
	2 Blue Group	1 Red Group	-3.80	1.796	.168	-8.64	1.04
		3 Green Group	1.70	1.796	.780	-3.14	6.54
		4 Yellow Group	3.80	1.796	.168	-1.04	8.64
	3 Green Group	1 Red Group	-5.50*	1.796	.021	-10.34	-.66
		2 Blue Group	-1.70	1.796	.780	-6.54	3.14
		4 Yellow Group	2.10	1.796	.650	-2.74	6.94
	4 Yellow Group	1 Red Group	-7.60*	1.796	.001	-12.44	-2.76
		2 Blue Group	-3.80	1.796	.168	-8.64	1.04
		3 Green Group	-2.10	1.796	.650	-6.94	2.74
Games-Howell	1 Red Group	2 Blue Group	3.80	2.145	.319	-2.27	9.87
		3 Green Group	5.50*	1.691	.024	.65	10.35
		4 Yellow Group	7.60*	1.720	.002	2.68	12.52
	2 Blue Group	1 Red Group	-3.80	2.145	.319	-9.87	2.27
		3 Green Group	1.70	1.869	.800	-3.71	7.11
		4 Yellow Group	3.80	1.896	.230	-1.67	9.27
	3 Green Group	1 Red Group	-5.50*	1.691	.024	-10.35	-.65
		2 Blue Group	-1.70	1.869	.800	-7.11	3.71
		4 Yellow Group	2.10	1.362	.435	-1.75	5.95
	4 Yellow Group	1 Red Group	-7.60*	1.720	.002	-12.52	-2.68
		2 Blue Group	-3.80	1.896	.230	-9.27	1.67
		3 Green Group	-2.10	1.362	.435	-5.95	1.75

\*. The mean difference is significant at the .05 level.

## Means Plots

