

SPSS PC Version 10: Analysis of Variance (ANOVA)¹

ANOVA involves the comparison of means from more than two groups. The ANOVA procedure results in an F-ratio that tells us the ratio of the variation between groups to the variation within groups. As is true in other hypothesis testing procedures, the purpose is to determine whether or not the differences you observe between the *samples* represents a *real difference* between the *populations* from which the samples were drawn or, alternatively, the observed difference just happened by chance when, in reality, the populations don't differ at all on the variable you are interested in.

Using SPSS to conduct an ANOVA test:

- To conduct an ANOVA test choose *Analyze, Compare Means, One-way ANOVA*. A dialog box will appear in which you must specify two things: the "dependent list" and the "Factor." The "Factor" is the variable that defines the groups that you want to compare. The "dependent list" will contain all of those variables on which you would like to compare the categories of your "Factors." For example, let's examine whether the average level of education among women varies across the regions of the country. Here our "dependent list" would contain the variable *educ* ("Education in years") and the "Factor" would be the variable *region* ("Region where R lives"). The syntax for this is:

```
ONEWAY  
educ BY region  
/MISSING ANALYSIS .
```

- The results of this test:

ANOVA

EDUC Education in Years

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	356.025	3	118.675	15.512	.000
Within Groups	82954.659	10843	7.651		
Total	83310.684	10846			

- The output from the ANOVA procedure is actually quite easy to read. In the main box of the output, appropriately labeled "ANOVA," are all of the main pieces of the ANOVA test. Included here are the "sum of squares between groups," the "sum of squares within groups," and the degrees of freedom associated with each. Also included are the "mean squares" between groups and within groups. These are obtained by dividing each "sum of squares" by the appropriate degrees of freedom. Finally, and most important, is the F-ratio and significance level associated with this F(obtained). The F-ratio is calculated by dividing the mean squares between groups by the mean squares within groups. If this value is large enough (larger than the critical value of F) and, therefore, the probability of obtaining an F-ratio of this size or larger (labeled "Sig." in the SPSS output) is small enough (smaller than our chosen alpha), then we can reject the null hypothesis in favor of the idea that the populations we are comparing really do differ from each other.
- In our example, the F(obtained) is 15.512 which far exceeds the F-critical of 2.60 for this test when using an alpha of .05. Correspondingly, the observed p-value of .000 is well below my chosen alpha of .05. By either standard, I can reject the null hypothesis in support of the conclusion that the level of female education really does vary significantly across the regions of the country. In fact, even if I had used a more stringent alpha level (e.g., .01 or .001) I would have still rejected the null hypothesis.
- Note that if you choose *Options* and the *Descriptive Statistics* in the main ANOVA dialog box, SPSS will also produce an additional box containing the means and standard deviations (and additional information, including Ns, minimums and maximums) for each group being contrasted in the ANOVA test. Although not essential, this is good information to have if you want to look for the source of any significant difference between means detected by the

¹Prepared by Kyle Crowder of the Sociology Department of Western Washington University, modified by Patty Glynn. November 24, 2000

ANOVA process. It may also be important to look at these to make sure that missing values are coded correctly, and that you do not have unexpected values because of dirt in your data. The syntax is:

```
ONEWAY
educ BY region
/STATISTICS DESCRIPTIVES
/MISSING ANALYSIS .
```

- You can also ask SPSS to give you information about what groups vary significantly from others by putting a check in the box **SHEFFEE** and/or **BONFERRONI** in the **POST HOC** box. Bonferroni is a stricter test, which takes into consideration the increased possibility of finding significant relationships because multiple tests are being performed. The syntax and results would look like:

```
ONEWAY
educ BY region
/STATISTICS DESCRIPTIVES
/MISSING ANALYSIS
/POSTHOC = SCHEFFE BONFERRONI ALPHA(.05) .
```

And the results, as follows, shows that there is a significant difference between those living in the Northeast and South and West, but not between those living in the Northeast and the Midwest.

Multiple Comparisons

Dependent Variable: Education in Years

(I) Region Where R Lives	(J) Region Where R Lives	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
Scheffe	Northeast	Midwest	4.354E-02	8.198E-02	.963	-.1857	.2728
		South	.4303*	7.622E-02	.000	.2171	.6434
		West	.2882*	8.284E-02	.007	5.662E-02	.5199
	Midwest	Northeast	-4.3543E-02	8.198E-02	.963	-.2728	.1857
		South	.3867*	7.064E-02	.000	.1892	.5842
		West	.2447*	7.774E-02	.019	2.736E-02	.4620
	South	Northeast	-.4303*	7.622E-02	.000	-.6434	-.2171
		Midwest	-.3867*	7.064E-02	.000	-.5842	-.1892
		West	-.1420	7.164E-02	.269	-.3423	5.828E-02
	West	Northeast	-.2882*	8.284E-02	.007	-.5199	-5.6619E-02
		Midwest	-.2447*	7.774E-02	.019	-.4620	-2.7356E-02
		South	.1420	7.164E-02	.269	-5.8277E-02	.3423
Bonferroni	Northeast	Midwest	4.354E-02	8.198E-02	1.000	-.1728	.2599
		South	.4303*	7.622E-02	.000	.2291	.6314
		West	.2882*	8.284E-02	.003	6.964E-02	.5068
	Midwest	Northeast	-4.3543E-02	8.198E-02	1.000	-.2599	.1728
		South	.3867*	7.064E-02	.000	.2003	.5731
		West	.2447*	7.774E-02	.010	3.957E-02	.4498
	South	Northeast	-.4303*	7.622E-02	.000	-.6314	-.2291
		Midwest	-.3867*	7.064E-02	.000	-.5731	-.2003
		West	-.1420	7.164E-02	.285	-.3311	4.702E-02
	West	Northeast	-.2882*	8.284E-02	.003	-.5068	-6.9640E-02
		Midwest	-.2447*	7.774E-02	.010	-.4498	-3.9574E-02
		South	.1420	7.164E-02	.285	-4.7017E-02	.3311

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