

Comparing Two or more than Two Groups

Comparing Means, Conducting T-Tests
and ANOVA



Chapter 9/1: Comparing Two or more than Two Groups

Cross tabulation is a useful way of exploring the relationship between variables that contain only a few categories. For example, we could compare how men and women feel about abortion. Here, our dependent variable (abortion) consists of only two categories—approve or disapprove.

But, what if we wanted to find out if the average age at birth of first child is younger for women than for men? Here, our dependent variable is a continuous variable consisting of many values. We could recode it so that it only had a few categories (e.g., under 20, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 and older), but that would result in the loss of a lot of information. A better way to do this would be to compare the mean age at birth of first child for men and women.

We're going to use the subset from the 2008 General Social Survey to answer this question. Click on "Analyze", point your mouse at "Compare Means", and then click on "Means".

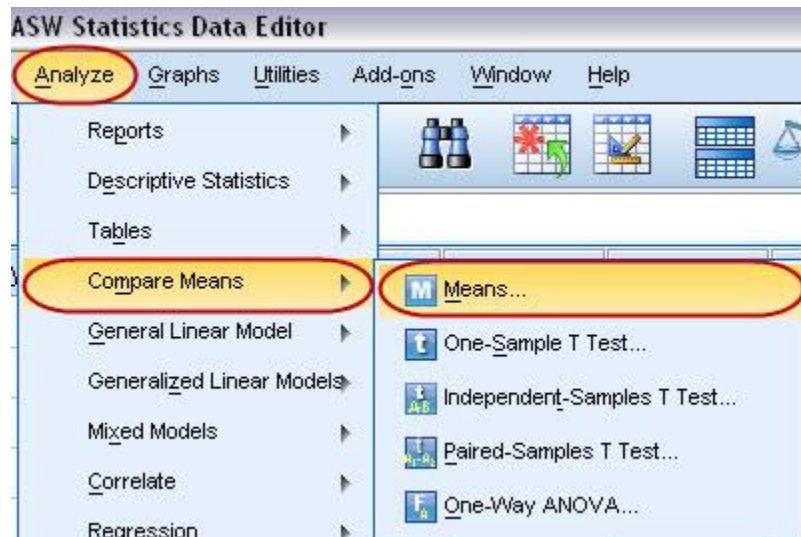


Figure 9- 1

We want to put age at birth of first child (**AGEKDBRN**) in the Dependent List and **SEX** in the Independent List. Highlight **AGEKDBRN** in the list of variables on the left of your screen, and then click on the arrow next to the Dependent List box. Now click on the list of variables on the left and use the scroll bar to find the variable **SEX**. Click on it to highlight it and then click on the arrow next to the Independent List box. Your screen should look like Figure 9-2. Click on "OK" and the Output Window should look like Table 9-1.

CRJ 716: Chapter 9 – Comparing Groups
The Existence, Strength, and Direction of an Association

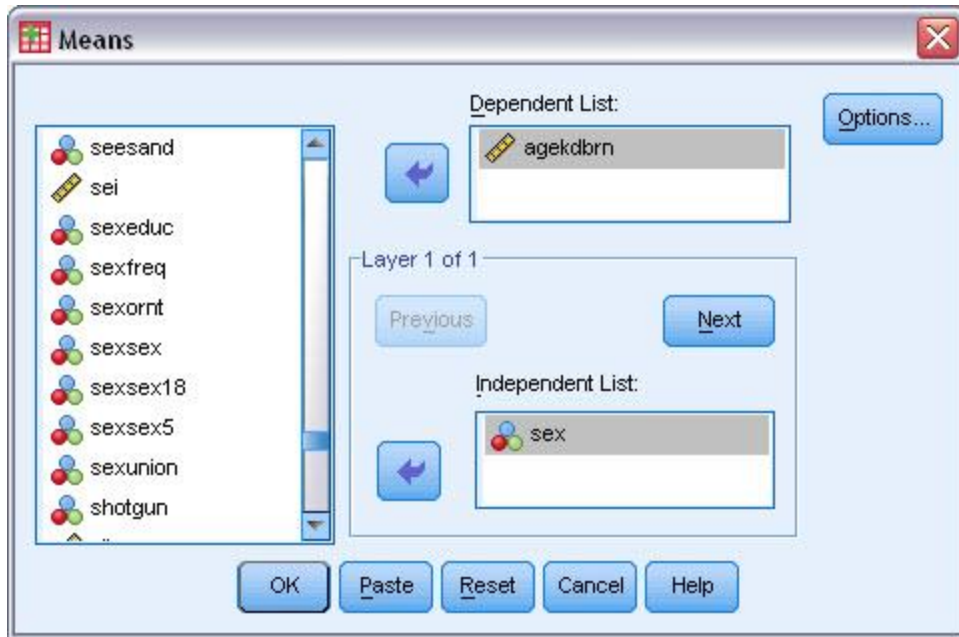


Figure 9- 2

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
R'S AGE WHEN 1ST CHILD BORN * RESPONDENTS SEX	1489	73.6%	534	26.4%	2023	100.0%

Report

R'S AGE WHEN 1ST CHILD BORN			
RESPONDENTS SEX	Mean	N	Std. Deviation
MALE	25.00	623	5.444
FEMALE	22.87	866	5.128
Total	23.76	1489	5.364

Table 9- 1

On the average, women are a little more than two years younger ($25 - 22.87 = 2.13$ years) than men at the birth of first child.

INDEPENDENT-SAMPLES T TEST

Independent groups of data contain measurements that pertain to two unrelated samples of items. Males and females are considered independent groups. Generally, for independent groups you perform the 2-sample *t-test*.

Let's take a real example from our GSS2008 data set. If married women are, on the average, two years younger than men at birth of first child, can we conclude that this is also true in our population? Can we make an inference about the population (all people) from our sample (about 2,023 people selected from the population)? To answer this question we need to do a *t-test*. This will test the null hypothesis that **men and women in the population do not differ in terms of their mean age at the birth of their first child**. The particular version of the *t-test* that we will be using is called the *independent-samples t test* since our two samples are completely independent of each other. In other words, the selection of cases in one of the samples does not influence the selection of cases in the other sample. We'll look later at a situation where this is not true.

We want to compare our sample of men with our sample of women and then use this information to make an inference about the population. Click on "Analyze", then point your mouse at "Compare Means" and then click on "Independent-Samples T Test".

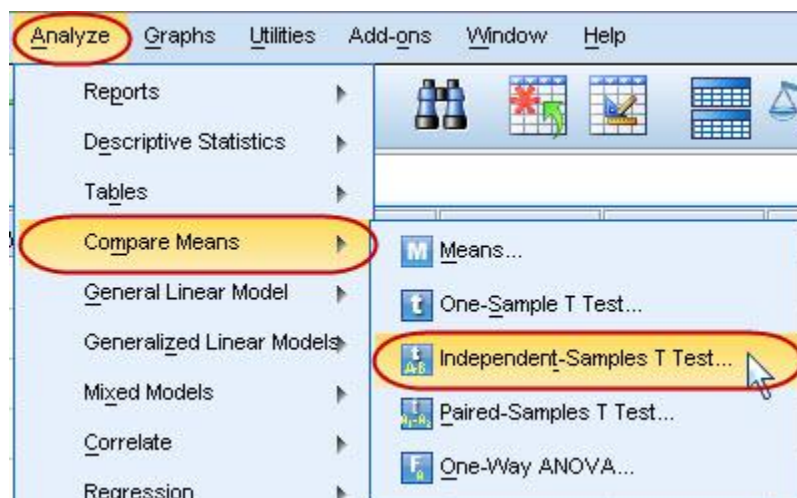


Figure 9- 3

Find **AGEKDBRN** in the list of variables on the left and click on it to highlight it, and then click on the arrow to the left of the Test Variable box. This is the variable we want to test so it will go in the Test Variable box. Now click on the list of variables on the left and use the scroll bar to find the variable **SEX**. Click on it to highlight it and then click on the arrow to the left of the Grouping Variable box. **SEX** defines the two groups we want to compare so it will go in the Grouping Variable box. Now we want to define the groups so click on the "Define Groups" button. This will open the Define Groups box. Since males are coded 1 and females 2, type 1 in the Group 1 box and 2 in the Group 2 box. (You will have to click in each box before typing the value.) This tells SPSS what the two groups are we want to compare. (If you don't know how males and females are coded, click on "Utilities" menu, then on "Variables" and scroll down until you find the variable **SEX** and click on it. The box to the right will tell you the values for males and females. Be sure to close this box.) Your screen should look like Figure 9-4.

CRJ 716: Chapter 9 – Comparing Groups
The Existence, Strength, and Direction of an Association

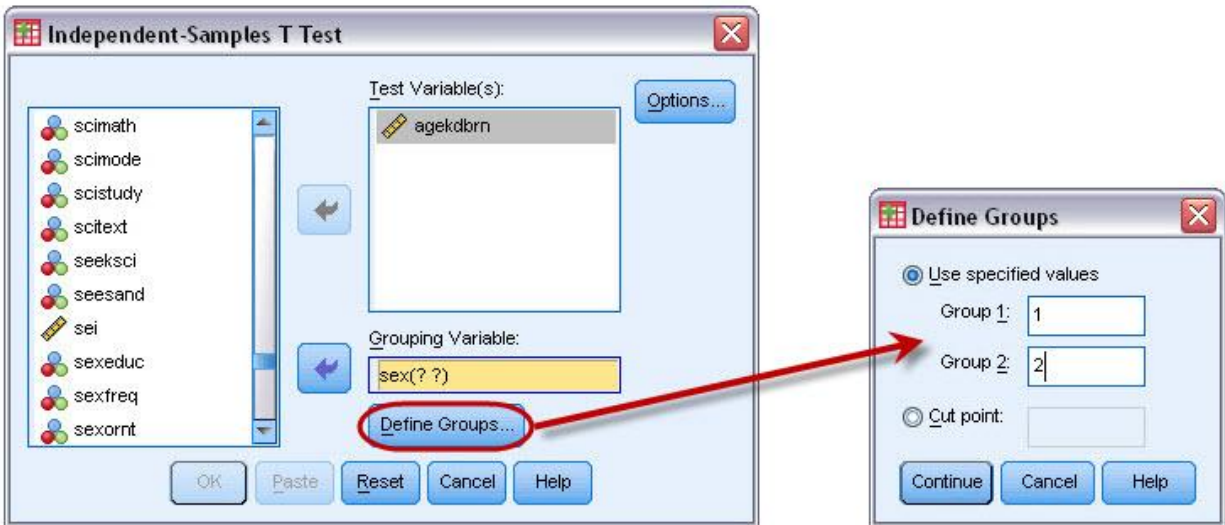


Figure 9- 4

Now click on "Continue" and on "OK" in the Independent-Samples T Test box.

Your output screen should look like Table 9-2.

T-Test										
Group Statistics										
	RESPONDENTS SEX	N	Mean	Std. Deviation	Std. Error Mean					
R'SAGE WHEN 1ST CHILD BORN	MALE	623	25.00	5.444	.218					
	FEMALE	866	22.87	5.128	.174					
Independent Samples Test										
Levene's Test for Equality of Variances										
t-test for Equality of Means										
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
R'SAGE WHEN 1ST CHILD BORN	Equal variances assumed	.724	.395	7.704	1487	.000	2.130	.276	1.588	2.672
	Equal variances not assumed			7.629	1291.1	.000	2.130	.279	1.582	2.677

Table 9- 2

This table shows you the mean age at birth of first child for men (25.00) and women (22.87) which is a mean difference of 2.13. It also shows you the results of two *t*-tests. Remember that this test's null hypothesis is that **men and women have the same mean age at the birth of their first child in the population**. There are two versions of this test. One assumes that the populations of men and women have equal variances (for AGEKDBRN), while the other doesn't make any assumption about the variances of the populations. The table also gives you the values for the degrees of freedom and the observed significance level. The significance value is **.000** for both versions of the *t*-test. Actually, this means less than .0005 since PASW rounds to the nearest third decimal place. This significance value is the probability that the *t*-value would be this big or bigger simply by chance if the null hypothesis was true. Since this probability is so small (less than five in 10,000), we will reject the null hypothesis and conclude that there probably is a statistical difference between men and women in terms of average age at the birth of their first child in the population. Notice that this is a two-tailed significance value. If you wanted the one-tailed significance value, just divide the two-tailed value in half.

CRJ 716: Chapter 9 – Comparing Groups The Existence, Strength, and Direction of an Association

Let's work another example. This time we will compare males and females in terms of average years of school completed (EDUC). Click on "Analyze", point your mouse at "Compare Means", and click on "Independent-Samples T Test". Click on "Reset" to get rid of the information you entered previously. Move EDUC into the Test Variable box and SEX into the Grouping Variable box. Click on "Define Groups" and define males and females as you did before. Click on "Continue" and then on "OK" to get the output window. Your screen should look like Table 9-3. There isn't much of a difference between men and women in terms of years of school completed (0.04 years), and from looking at the Sig. (2-tailed), we fail to reject the null hypothesis since the observed significance level is greater than .05 ($p\text{-value} = 0.772$). Remember that this test's null hypothesis is that **men and women have the same education level**. When we fail to reject the null hypothesis, we conclude that there is no statistical difference in education of men and women.

T-Test										
Group Statistics										
	RESPONDENTS SEX	N	Mean	Std. Deviation	Std. Error Mean					
HIGHEST YEAR OF	MALE	928	13.45	3.121	.102					
SCHOOL COMPLETED	FEMALE	1090	13.41	3.044	.092					

Independent Samples Test										
Levene's Test for Equality of Variances										
t-test for Equality of Means										
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
HIGHEST YEAR OF	Equal variances	.966	.326	.290	2016	.772	.040	.138	-.230	.310
SCHOOL COMPLETED	assumed									
	Equal variances not assumed			.290	1948.562	.772	.040	.138	-.230	.310

Table 9- 3

PAIRED-SAMPLES T TEST

We said we would look at an example where the samples are not independent. (SPSS calls these **paired samples**. Sometimes they are called matched samples.) *Paired groups* contain measurements for one sample of items, but they contain two measurements for each item. A common example of paired groups is "before-and-after" measurement, where the goal of analysis is to decide if the average change from "before" to "after" is greater than it could happen by chance alone. For example, a doctor weighs thirty people before they begin a program to quit smoking and weighs them again six months after they have completed the program. The goal of analysis is to decide if the average weight change is greater than would be expected by chance.

Back to our data set... Let's say we wanted to compare the educational level of the respondent's father and mother. **PAEDUC** is the years of school completed by the father and **MAEDUC** is years of school for the mother. Clearly our samples of fathers and mothers are not independent of each other. If the respondent's father is in one sample, then his or her mother will be in the other sample. One sample determines the other sample. Another example of paired samples is before and after measurements. We might have a person's weight before they started to exercise and their weight after exercising for two months. Since both measures are for the same person we clearly do not have independent samples. This requires a different type of *t-test* for paired samples.

Click on "Analyze", then point your mouse at "Compare Means", and then click on "**Paired-Samples T Test**". Scroll down to MAEDUC in the list of variables on the left and click on it to move it to the Current Selections box as Variable 1. Now click on PAEDUC to move it to the Current Selections box as Variable

CRJ 716: Chapter 9 – Comparing Groups The Existence, Strength, and Direction of an Association

2. Click on the arrow to the left of the Paired Variables box to move this pair of variables into the box in the middle of the window. Your screen should look like Figure 9-5.

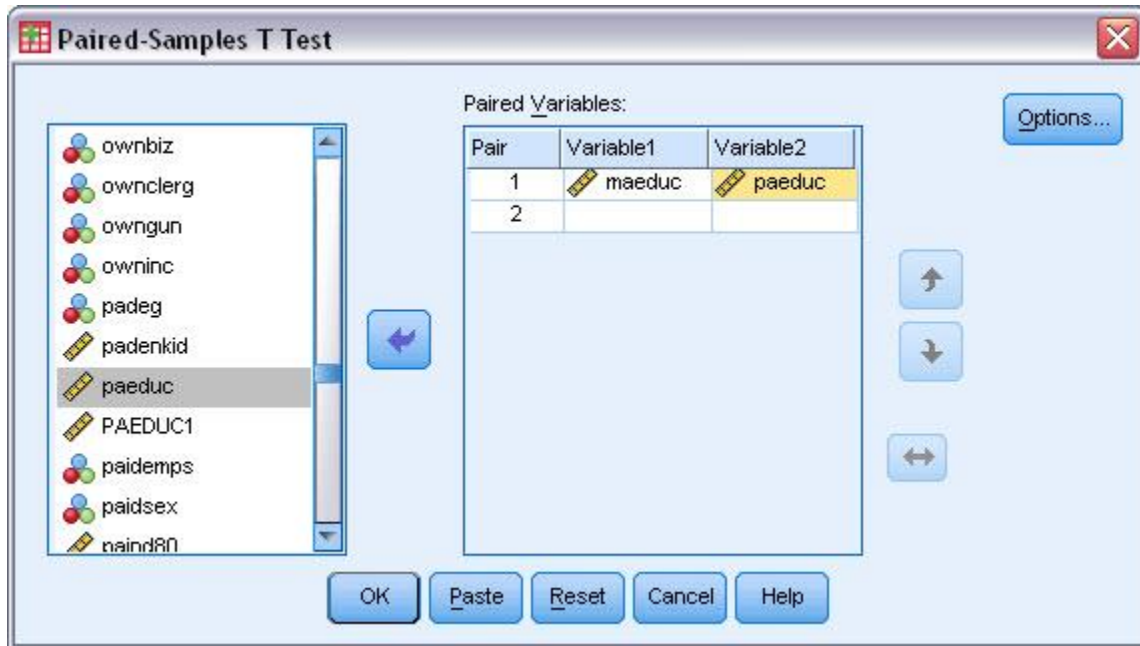


Figure 9- 5

Click on "OK" and your screen should look like Table 9-4.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HIGHEST YEAR SCHOOL COMPLETED, MOTHER	11.52	1397	3.703	.099
	HIGHEST YEAR SCHOOL COMPLETED, FATHER	11.53	1397	4.199	.112

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	HIGHEST YEAR SCHOOL COMPLETED, MOTHER & HIGHEST YEAR SCHOOL COMPLETED, FATHER	1397	.679	.000

Paired Samples Test									
Paired Differences									
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	HIGHEST YEAR SCHOOL COMPLETED, MOTHER - HIGHEST YEAR SCHOOL COMPLETED, FATHER	-.002	3.199	.086	-.170	.166	-.025	1396	.980

Table 9- 4

This table shows the mean years of school completed by mothers (11.52) and by fathers (11.53), as well as the standard deviations. The *t-value* for the paired-samples *t-test* is -0.025 and the 2-tailed significance value is 0.980. (You may have to scroll down to see these values.) This is the probability of

CRJ 716: Chapter 9 – Comparing Groups The Existence, Strength, and Direction of an Association

getting a t-value this large or larger just by chance if the null hypothesis is true. Since this probability is greater than .05, we will not reject the null hypothesis. There is no statistical basis for saying that the respondents' fathers and mothers have different educational levels. Actually, even if you see with your bare eyes, the mean difference of education levels is a tiny 0.01 (11.53-11.52).

ONE-WAY ANALYSIS OF VARIANCE

In this chapter we have compared two groups (males and females). What if we wanted to compare more than two groups? For example, we might want to see if age at the birth of their first child (AGEKDBRN) varies by educational level. This time let's use the respondent's highest degree (DEGREE) as our measure of education. To do this we will use One-Way Analysis of Variance (often abbreviated **ANOVA**).

Click on "Analyze", then point your mouse at "Compare Means", and then click on "Means". Click on "Reset" to get rid of what is already in the box. Click on AGEKDBRN to highlight it and then move it to the Dependent List box by clicking on the arrow to the left of the box. Then scroll down the list of variables on the left and find DEGREE. Click on it to highlight it and move it to the Independent List box by clicking on the arrow to the left of this box. Click on the "Options" button and this will open the Means: Options box. Click on the box labeled "Anova table and eta". This should put a check mark in this box indicating that you want SPSS to do a One-Way Analysis of Variance. Your screen should look like Figure 9-6. Click on "Continue" and then on "OK" in the Means box and your screen should look like Table 9-5 on the next page.

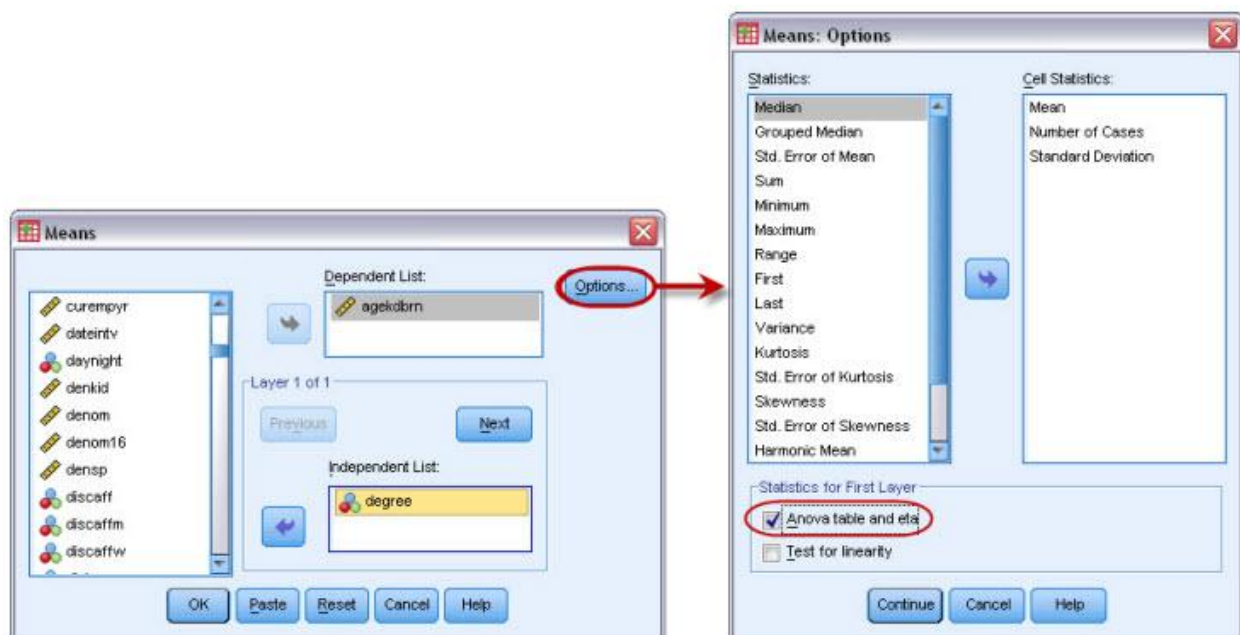


Figure 9- 6

CRJ 716: Chapter 9 – Comparing Groups
The Existence, Strength, and Direction of an Association

Report			
R'S AGE WHEN 1ST CHILD BORN			
RS HIGHEST DEGREE	Mean	N	Std. Deviation
LT HIGH SCHOOL	21.04	237	4.774
HIGH SCHOOL	22.72	753	4.686
JUNIOR COLLEGE	23.67	126	4.706
BACHELOR	27.31	237	4.996
GRADUATE	28.21	136	5.456
Total	23.76	1489	5.364

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
R'S AGE WHEN 1ST CHILD BORN * RS HIGHEST DEGREE	Between Groups (Combined)		8251.424	4	2062.856	88.558	.000
	Within Groups		34567.890	1484	23.294		
	Total		42819.314	1488			

Measures of Association		
	Eta	Eta Squared
R'S AGE WHEN 1ST CHILD BORN * RS HIGHEST DEGREE	.439	.193

Table 9- 5

In this example, the independent variable has five categories: less than high school, high school, junior college, bachelor, and graduate. Table 9-5 shows the mean age at birth of first child for each of these groups and their standard deviations, as well as the Analysis of Variance table including the sum of squares, degrees of freedom, mean squares, the F-value and the observed significance value. (You will have to scroll down to see the Analysis of Variance table.) The significance value for this example is the probability of getting a F-value of 88.558 or higher if the null hypothesis is true. Here the null hypothesis is that **the mean age at birth of first child is the same for all five population groups**. In other words, that the mean age at birth of first child for all people with less than a high school degree is equal to the mean age for all with a high school degree and all those with a junior college degree and all those with a bachelor's degree and all those with a graduate degree. Since this probability is so low (.000 or less than 5% [<0.05]), we would reject the null hypothesis and conclude that these population means are probably not all the same or that the difference of the mean age at first child is statistically significant among the categories.

There is another procedure in SPSS that does One-Way Analysis of Variance and this is called One-Way ANOVA. This procedure allows you to use several multiple comparison procedures that can be used to determine which groups have means that are significantly different. If you want to use these procedures, consult the SPSS 11.0 Syntax Reference Guide.

Summary

This chapter has explored ways to compare the means of two or more groups and statistical tests to determine if these means differ significantly. These procedures would be useful if your dependent variable was continuous and your independent variable contained a few categories. The next chapter looks at ways to explore the relationship between pairs of variables that are both continuous.