CHAPTER 4

PEARSON'S *R*, CHI-SQUARE, *T*-TEST, AND ANOVA

OUTLINE OF CHAPTER

SECTION 1: INTRODUCTION AND THEORETICAL BACKGROUND

Introduction

Inferential statistics consist of statistical methods that are used to test hypotheses that relate to relationships between variables. For example, you might hypothesize that individuals with greater levels of education tend to have higher incomes. While we can use descriptive statistics such as line plots as outlined in the previous chapter to illustrate the relationship between these two variables, we need to use inferential statistics to more rigorously demonstrate whether or not there is a relationship between these two variables.

With all inferential statistics, which particular statistical test you use will depend on the nature of your data as well as the nature of your hypothesis. In this chapter, **Pearson's correlation coefficient** (also known as **Pearson's *r***), the chi-square test, the *t*-test, and the ANOVA will be covered. Pearson's correlation coefficient (r) is used to demonstrate whether two variables are correlated or related to each other. When using Pearson's correlation coefficient, the two variables in question must be continuous, not categorical. So it can be used, for example, to test the relationship between years of education and income, as these are both continuous variables, but not race and highest degree completed, as these are categorical variables. The **chi-square statistic** is used to show whether or not there is a relationship between two categorical variables. For example, you can use the chi-square statistic to show the relationship between the highest degree completed (e.g., coded as none, high school diploma, bachelors, etc.) and political affiliation (coded as Republican or Democrat). The *t*-test is used to test whether there is a difference between two groups on a continuous dependent variable. For example, you would select the *t*-test when testing whether there is a difference in income between males and females. The ANOVA is very similar to the *t*-test, but it is used to test differences between three or more groups. For example, you would use an ANOVA to test whether there is a difference in income between blacks, whites, and Hispanics. The ANOVA is actually a generalized form of the *t*-test, and when conducting comparisons on two groups, an ANOVA will give you identical results to a *t*-test.

Pearson's r: Theory

The purpose of the correlation coefficient is to determine whether there is a significant relationship (i.e., correlation) between two variables. The most commonly used correlation coefficient is the one published by Karl Pearson in 1895, having been developed earlier by Sir Francis Galton. It goes under several names, including Pearson's r, the product-moment correlation coefficient, and Pearson's correlation coefficient. I will typically refer to it as Pearson's *r* for the sake of brevity.

Pearson's *r* is used to illustrate the relationship between two continuous variables, such as years of education completed and income. The correlation between any two variables using Pearson's *r* will always be between –1 and +1. A correlation coefficient of 0 means that there is no relationship, either positive or negative, between these two variables. A correlation coefficient of +1 means that there is a perfect positive correlation, or relationship, between these two variables. In the case of +1, as one variable increases, the second variable increases in exactly the same level or proportion. Likewise, as one variable decreases, the second variable would decrease in exactly the same level or proportion. A correlation coefficient of –1 means that there is a perfect negative correlation, or relationship, between two variables. In this case, as one variable increases, the second variable decreases in exactly the same level or proportion. Also, as one variable decreases, the other would increase in exactly the same level or proportion.

You most likely will never see a correlation between two variables of –1 or +1 in the social sciences as while two variables may be very highly related, the chance of error or random variation is too great to have a perfect correlation. A positive correlation means that generally, as one variable increases, the other will increase, and as one variable decreases, the other will decrease. Also, a negative correlation means that in general, if one variable increases, the other will decrease, and as one variable decreases, the other will increase. Very important here is the notion of significance, which I introduced you to in Chapter 1. When determining Pearson's *r*, or other correlation coefficients, it is important to be aware of whether your correlation is in fact significant or not at the .05 level.

Let's now compute the Pearson's *r* for some data. The table below consists of data made up for this example.

Years of Education (x)	Income (in Thousands of $) (y)
8	12
12	15
8	8
14	20
12	18
16	45
20	65
24	85
24	100
24	90

As you may have noticed, I tried to create a positive relationship between years of education and income—I am hoping that this will result in a strong positive correlation coefficient that will be significant.

The equation for Pearson's r is as follows:

$$r = \frac{\sum xy - N\bar{x}\bar{y}}{\sqrt{(\sum x^2 - N\bar{x}^2)(\sum y^2 - N\bar{y}^2)}}$$

This equation requires us to first calculate the sum of the product of all our data pairs, the means of both variables, and the sum of the squared values of both variables.

So first,

$$\begin{aligned}\sum xy &= (8 \times 12) + (12 \times 15) + (8 \times 8) + (14 \times 20) + (12 \times 18) + (16 \times 45) \\ &\quad + (20 \times 65) + (24 \times 85) + (24 \times 100) + (24 \times 90) \\ &= 96 + 180 + 64 + 280 + 216 + 720 + 1{,}300 + 2{,}040 + 2{,}400 + 2{,}160 \\ &= 9{,}456\end{aligned}$$

$$\bar{x} = \frac{8 + 12 + 8 + 14 + 12 + 16 + 20 + 24 + 24 + 24}{10} = \frac{162}{10} = 16.2$$

$$\bar{y} = \frac{12 + 15 + 8 + 20 + 18 + 45 + 65 + 85 + 100 + 90}{10} = \frac{458}{10} = 45.8$$

$$\sum x^2 = 8^2 + 12^2 + \text{L} + 24^2 = 2{,}996$$

$$\sum y^2 = 12^2 + 15^2 + \text{L} + 90^2 = 32{,}732$$

N = Number of cases or data pairs = 10.

Now, plugging these values into our equation, we get the following:

$$\begin{aligned}r &= \frac{\sum xy - N\bar{x}\bar{y}}{\sqrt{(\sum x^2 - N\bar{x}^2)(\sum y^2 - N\bar{y}^2)}} \\ &= \frac{9456 - 10(16.2)(45.8)}{\sqrt{(2996 - 10(16.2^2))(32732 - 10(45.8^2))}} \\ &= \frac{9456 - 7419.6}{\sqrt{(2996 - 2624.4)(32732 - 20976.4)}} = \frac{2036.4}{\sqrt{4368380.96}} = 0.9743\end{aligned}$$

I will use this same example in the sections on IBM SPSS and Stata—in those sections, you will be able to see that the result for Pearson's r using either of these programs is identical to the value we have calculated by hand.

Now, we can see that our correlation, .9743, is very high as it is very close to +1, the maximum possible value for Pearson's r. But we still need to calculate the p value in order to determine whether this correlation is statistically significant or not.

To determine this, we will first calculate a t ratio using the following equation:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Now, plugging our values into the equation, we get the following:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} = \frac{.9743\sqrt{10-2}}{\sqrt{1-.9743^2}} = \frac{.9743\sqrt{8}}{\sqrt{.0507}} = \frac{2.7557}{.2251} = 12.2386$$

Also, we will need to know our degrees of freedom (*df*). This is equal to the number of pairs of data minus 2:

$$df = N - 2 = 10 - 2 = 8$$

Next, we will need to consult a t table to compare our calculated t value with the critical t value in order to determine statistical significance. Looking at a t table, we can see that for 8 degrees of freedom, the critical t value for a p level of .05 (two-tailed) is 2.306. As our calculated t value is greater than the critical t value at the .05 level, we can say that the correlation between education and income is significant at the .05 level. Again referring to our table, we can see that our correlation is even significant at the .001 level, as the critical t value in this case is 5.041, which is still lower than our calculated t value. This means that the probability that the correlation between education and income is simply due to error or chance is less than 0.1%. In this example, I have used the two-tailed critical t value, which is more conservative than a one-tailed test and is generally preferred. If you are not making a **directional hypothesis** (examples of a directional hypothesis: those with greater levels of education will have higher incomes or males have higher incomes than females), then you would use a **two-tailed test**, as it does not make any specification regarding direction. For example, a two-tailed test would be used if you're simply hypothesizing that there will be a correlation between level of education and income, but not specifying the direction of the correlation. However, if you were making a directional hypothesis, for example that those with more education are more likely to have higher incomes, the **one-tailed test** could be used. However, when the direction between your two variables corresponds to the direction stated in your hypothesis, the one-tailed test is less conservative than the two-tailed test and so tends to be used less often.

In the next section, the concept of ***R*-squared** will be discussed. The *R*-squared value represents the proportion of variance in the dependent variable (the variable you are trying to predict or explain) that is explained by the independent variable(s) (the variables that you are using to explain or predict the dependent variable). In this example, it would make sense that we would use years of education to predict the respondent's income and not vice versa. What's interesting is that we simply need to square the value we arrived at after calculating Pearson's *r* to attain the *R*-squared. Thus,

$$R^2 = r^2 = .9743^2 = .9493$$

Later on, in the Stata section, I will replicate this result. We can interpret this by stating that level of education explains 94.93% of the variance in income. Here, I simply moved the decimal point two places to the right to arrive at this value.

Finally, it is important to state again that Pearson's *r* is only used for continuous variables. To determine the correlation between variables that are ordered and categorical or dichotomous, there are a number of special options, including **Kendall's tau**, **Spearman's rank correlation coefficient** or **Spearman's rho**, the **polyserial correlation**, the **polychoric correlation**, **phi**, the **tetrachoric correlation**, and others. Many of these tests require specialized software programs or certain specific add-ons to IBM SPSS or Stata. These additional measures of correlation are described in more detail in Appendix C, Section 4, Part F.

Chi-Square: Theory

The chi-square statistic is used to show whether or not there is a relationship between two categorical variables. It can also be used to test whether or not a number of outcomes are occurring in equal frequencies or not, or conform to a known distribution. For example, when rolling a die, there are six possible outcomes. After rolling a die hundreds of times, you could tabulate the number of times each outcome occurred and use the chi-square statistic to test whether these outcomes were occurring in basically equal frequencies or not (e.g., to test whether the die is weighted). The chi-square statistic was also developed by Karl Pearson.

This is the chi-square equation:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Here,

χ^2 = the chi-square statistic

O_i = the observed frequency

E_i = the expected frequency

i = the number of the cell (cell 1, cell 2, etc.)

Here, the summation is simple. We simply calculate the square of the difference between the observed and expected frequency and divide that by the expected frequency for each cell. Then, we simply sum all these quotients together. The concept of a "cell" is also easy to understand. If we are testing whether a number of outcomes are occurring in equal frequencies or not, such as in the example of the die, we would count each outcome as a cell. If we were testing a relationship between two variables, say between degree and political affiliation, the data would look like this:

	Political Affiliation	
Degree	**Republican**	**Democrat**
None	23	45
HS	17	42
BA	28	35
MA	32	32
Above MA	42	28

And we would have 10 cells all together.

Now, let's use these two examples to calculate the chi-square statistic. Say we roll a die 600 times and get the following results:

Outcome	Frequency
1	95
2	72
3	103
4	105
5	97
6	128

Here, we want to calculate the chi-square statistic to see whether these outcomes are occurring at basically the same frequencies or not. Now, if you remember from previous chapters, simply because the numbers are not exactly the same does not necessarily mean that certain outcomes are occurring more frequently than others in the statistical sense. To find out whether this is true, we need to run a statistical test and find the probability value (*p* value). To calculate the chi-square value for this particular example, we need to simply plug these numbers into the equation, as shown below. One hundred is chosen as the expected value for all cells, as it would be expected that you would get an equal number of each outcome (100 1s, 100 2s, etc.).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$
$$= \frac{(95-100)^2}{100} + \frac{(72-100)^2}{100} + \frac{(103-100)^2}{100} + \frac{(105-100)^2}{100} + \frac{(97-100)^2}{100} + \frac{(128-100)^2}{100}$$
$$= \frac{25}{100} + \frac{784}{100} + \frac{9}{100} + \frac{25}{100} + \frac{9}{100} + \frac{784}{100} = 16.36$$

So 16.36 is our chi-square statistic for this example, but we still do not know whether or not this value is significant (i.e., if the probability level is below .05 or not). To do this next step, you need to calculate the degrees of freedom. In this example, and in all examples in which we are simply looking at the frequencies of the different responses for single variable, degrees of freedom simply equals the number of different responses minus one. So we get,

Degrees of freedom = Number of response categories – 1 = 6 – 1 = 5

Now that we know both the chi-square value and the degrees of freedom, we simply need to look at a chi-square table to find the critical chi-square value for our degrees of freedom using the .05 probability level.

	Probability Level		
Degrees of Freedom	**.05**	**.01**	**.001**
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.34	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
⋮			

So when looking at this table, we will move down to 5 degrees of freedom, and look at the first column specified by the .05 probability level. Here, we can see that the critical chi-square value for our example is 11.07. We calculated a chi-square value of 16.36. Since the chi-square value that we calculated is greater than the critical chi-square value for the .05 probability level, our results are statistically significant. This means that the die appears to be not being rolled fairly, that some outcomes occur more frequently than others, and that this difference is statistically significant at the .05 level. Looking again at our chi-square table, we can see that our calculated value is also greater than the critical chi-square value at the .01 probability level at 5 degrees of freedom. This means that our results are also significant at the more stringent .01 probability level (meaning that there is a less than 1% chance that these differences between outcomes are not actually significantly different and are instead due to error or chance).

Next, we will calculate the chi-square statistic using the example of political affiliation and the highest degree completed. Here, the equation for the chi-square statistic remains the same. However, degrees of freedom are calculated differently than before. In the case where there are two variables, degrees of freedom are calculated using this equation:

$$\text{Degrees of freedom} = (\text{Rows} - 1)(\text{Columns} - 1)$$

Here is a reproduction of the table from the previous page:

	Political Affiliation		
Degree	**Republican**	**Democrat**	**Total**
None	23	45	68
HS	17	42	59
BA	28	35	63
MA	32	32	64
Above MA	42	28	70
Total	142	182	324

So to calculate the chi-square statistic for this example, we need to do the following.

First, we need to calculate the expected values. In the example with the die, we do not need to formally calculate expected values, since there were six possible outcomes with equal probabilities. We simply divided 600 (the number of times we rolled the die, or the number of cases) by the number of possible outcomes (6) to get 100, the expected value for each possible outcome.

When calculating the chi-square statistic between two variables, we use the following equation to determine the expected value for each cell:

$$E_i = \frac{(\text{Row total})(\text{Column total})}{\text{Grand total}}$$

For example, this is how you would calculate the expected value for the first cell in the top left corner (individuals with no degree who are Republican):

$$E_i = \frac{(\text{Row total})(\text{Column total})}{\text{Grand total}} = \frac{(68)(142)}{324} = \frac{9656}{324} = 29.80$$

So after calculating the expected value for each cell, we would plug all our numbers into the equation for the chi-square statistic:

$$\begin{aligned}
\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} &= \frac{(23-29.80)^2}{29.80} + \frac{(17-25.86)^2}{25.86} + \frac{(28-27.61)^2}{27.61} \\
&+ \frac{(32-28.05)^2}{28.05} + \frac{(42-30.68)^2}{30.68} \\
&+ \frac{(45-38.20)^2}{38.20} + \frac{(42-33.14)^2}{33.14} + \frac{(35-35.39)^2}{35.39} \\
&+ \frac{(32-35.95)^2}{35.95} + \frac{(27-39.32)^2}{39.32} \\
&= \frac{46.24}{29.80} + \frac{78.50}{25.86} + \frac{0.15}{27.61} + \frac{15.60}{28.05} + \frac{128.14}{30.68} + \frac{46.24}{38.20} \\
&+ \frac{78.50}{33.14} + \frac{0.15}{35.39} + \frac{15.60}{35.95} + \frac{151.78}{39.32} = 17.20
\end{aligned}$$

Now, we need to calculate the degrees of freedom. In cases where we are calculating the chi-square statistic between two variables, this is the equation that we use:

$$\text{Degrees of freedom} = (\text{Rows} - 1)(\text{Columns} - 1)$$

So in our example, this would be our degrees of freedom:

$$df = (5-1)(2-1) = 4$$

So now we know that our chi-square value is 17.20 and our degrees of freedom is 4. Looking at our chi-square table, we see that the critical chi-square value for 4 degrees of freedom at the .05 probability level is 9.49. Since our calculated chi-square value is greater than the critical chi-square value, our results are significant at the .05 probability level. We can also see that our results are also significant at the .01 probability level, but not at the .001 probability level. Therefore, there is a statistically significant relationship between highest degree completed and political affiliation using either .05 or .01 as our standard.

t-Test: Theory

As mentioned in the introduction to this chapter, *t*-tests are used when you want to test the difference between two groups on some continuous variable. A good example here would be the difference in yearly income between males and females. *t*-tests can also be used when testing the same group of people at two different times; for example, testing whether there was a significant increase or decrease in the test scores of the same group of students at two different times.

The equation for the *t*-test depends on whether we are doing an ***independent samples* t*-test*** (comparing two different groups) or a **dependent samples *t*-test**, also called a ***paired* t*-test*** (comparing the same group at two different periods of time, or two different groups that have been "matched" on some important variable). There is also a **one-sample *t*-test** that is used to compare a group of scores with a known population mean. Furthermore, there are separate equations for the *independent samples t-test* depending on whether or not our two groups have equal sample sizes.

This is the equation for a one-sample *t*-test:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where

t = the t statistic

$\bar{x}$ = the mean of the sample

μ = the comparison mean

s = the sample standard deviation

n = the sample size

A *t*-test would be preferred to a *z*-test in situations where the sample size is less than 30, and the population standard deviation is unknown. If either the sample is greater than 30, *OR* the population standard deviation is known, you would prefer the *z*-test, which is covered in Appendix C, Section 4, Part A.

Say we had a sample of 10 individuals who had all taken an exam. If we wanted to test whether their scores, all together, are significantly different from the score of 100, we could use a one-sample *t*-test. First, we would calculate the mean of the sample and the sample standard deviation, both of which were covered in the previous chapter. Say that the mean of scores for these 10 individuals is 107.8, and the standard deviation is 5.35. To calculate the *t* statistic, we would simply plug these values into the equation:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{107.8 - 100}{5.35/\sqrt{10}} = 4.61$$

In this example, we have selected 100 as the value for the comparison mean as we want to test whether the scores in our sample significantly differ from 100. If we wanted to, we could test whether the scores were significantly different from another value, such as 110, by simply plugging this value in for the comparison mean.

Next, we need to calculate the degrees of freedom. Here, the degrees of freedom is simply the sample size minus one. Therefore,

$$\text{Degrees of freedom} = n - 1 = 10 - 1 = 9$$

Now, we will refer to a *t* table to determine the critical *t* value for 9 degrees of freedom at the .05 level of significance. Looking at a *t* table, this value is 2.26 (two-tailed *t*-test). Since our calculated *t* value of 4.61 is greater than the critical *t* value of 2.26, we can say that the scores of our sample of 10 individuals differ significantly from the score of 100. This effect is statistically significant at the .05 probability level. The *t* value for 9 degrees of freedom at the .01 level of significance is 3.25, while the *t* value for 9 degrees of freedom at the .001 level of significance is 4.78 (both two-tailed). Since our calculated *t* statistic of 4.61 is greater than the critical *t* value for the .01 level of significance, we can say that our result is statistically significant at the .01 probability level. As mentioned previously, when writing up results, you will mention only the most strict level of significance that you're able to obtain, whether .05, .01, or .001. In this case, we would mention only the .01 probability level in our results. For example, we could say the following: Our sample's mean of 107.8 was significantly different from 100 ($t = 4.61$, $df = 9$, $p < .01$).

This is the equation for the independent samples *t*-test when you have unequal sample sizes:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1} + \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2}}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Here,

$\bar{X}_1$ and $\bar{X}_2$ are the means of the two different groups

$n_1 = n$ of Group 1

$n_2 = n$ of Group 2

SS = sum of squares

Say we had two classes, one with five students and the other with seven students.

These were their scores:

Case	Group 1	Group 2
1	78	87
2	82	92
3	87	86
4	65	95
5	75	73
6	82	
7	71	

First we would calculate the means of each group. The mean (average) of Group 1 is 77.14, and the mean for Group 2 is 86.60.

Next, we calculate the sum of squares (SS) for each group. As you can see from the above equation,

$$SS = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

So for Group 1,

$$SS_1 = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1} = \left(78^2 + 82^2 + 87^2 + 65^2 + 75^2 + 82^2 + 71^2\right)$$
$$- \frac{(78 + 82 + 87 + 65 + 75 + 82 + 71)^2}{7}$$
$$= 41992 - \frac{540^2}{7} = 334.86$$

And, for Group 2,

$$SS_2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} = \left(87^2 + 92^2 + 86^2 + 95^2 + 73^2\right) - \frac{(87 + 92 + 86 + 95 + 73)^2}{5}$$
$$= 37783 - \frac{433^2}{5} = 285.20$$

Finally, plugging all these values into the t-test equation, we get the following:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{77.14 - 86.60}{\sqrt{\left[\frac{334.86 + 285.20}{7 + 5 - 2}\right]\left[\frac{1}{7} + \frac{1}{5}\right]}}$$
$$= \frac{-9.46}{\sqrt{\left(\frac{620.06}{10}\right)\left(\frac{12}{35}\right)}} = \frac{-9.46}{\sqrt{21.26}} = -0.44$$

Now, to see whether this is significant or not, we need to do the same thing as we did after calculating the chi-square statistic: Compare this value to the critical t value from a t table. First, we need to get the degrees of freedom.

For an independent, or between-subjects, t-test,

$$df = n_1 + n_2 - 2$$

which means, in our example, we have 10 degrees of freedom.

Here is a truncated t table:

	Two-Tailed t-Test: p Level		
df	**.05**	**.01**	**.001**
1	12.706	63.657	636.619
2	4.303	9.925	31.598
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587

And the top of this table I mention that these critical t scores are for the two-tailed t-test. The two-tailed t-test is used when you are not hypothesizing a direction in the relationship between your two groups and the dependent variable. For example, if you're testing the relationship between gender and religious attendance, and do not have a hypothesis, you would use the critical t scores from a two-tailed t-test table or column. The one-tailed t-test *can* be used if you are hypothesizing a directional relationship, for example, if you are hypothesizing that males will have higher incomes than females or that females will have greater religious attendance than males. However, the two-tailed t-test is a more stringent test and tends to be preferred over the one-tailed t-test, regardless of whether or not you have a directional hypothesis. This is true not only in regard to t-tests specifically but in general.

So in this example, we calculated a t score of -0.44. Before making the comparison with our critical t score table, we can first take the absolute value of this, which is 0.44 (i.e., simply make this number positive if it is a negative number). Now, for the .05 probability level with 10 degrees of freedom, we see from our table that the critical t score is 2.228 for a two-tailed test. Since our calculated t score is lower than the critical t score, our results are not significant at the .05 probability level. So the differences in the means of the scores that we saw between the two groups cannot be statistically attributed to any meaningful difference between these two groups. Here, if we wanted to report this result, we could simply say the following: The differences in test scores between our two groups were not statistically significant at the .05 probability level.

When we are performing an independent samples t-test (between subjects) for two groups having equal sample sizes (n), our equation can be simplified like this:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{2n - 2}\right]\left[\frac{2}{n}\right]}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2(SS_1 + SS_2)}{2n^2 - 2n}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2(SS_1 + SS_2)}{2(n^2 - n)}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{SS_1 + SS_2}{n^2 - n}}}$$

where n is the sample size of either group.

For example, say we have the following two groups of scores:

Case	Group 1	Group 2
1	63	88
2	57	95
3	48	84
4	52	99
5	38	87
Mean	51.6	90.6

First we would find the sum of squares for each group:

$$SS_1 = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1} = \left(63^2 + 57^2 + 48^2 + 52^2 + 38^2\right) - \frac{(63 + 57 + 48 + 52 + 38)^2}{5}$$

$$= 13670 - \frac{258^2}{5} = 357.20$$

$$SS_2 = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2} = \left(88^2 + 95^2 + 84^2 + 99^2 + 87^2\right) - \frac{(88 + 95 + 84 + 99 + 87)^2}{5}$$

$$= 41195 - \frac{453^2}{5} = 153.20$$

The more complex equation gives us the following:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{51.6 - 90.6}{\sqrt{\left[\frac{357.2 + 153.2}{5 + 5 - 2}\right]\left[\frac{1}{5} + \frac{1}{5}\right]}}$$

$$= \frac{-39}{\sqrt{\left(\frac{510.4}{8}\right)\left(\frac{2}{5}\right)}} = \frac{-39}{\sqrt{25.52}} = -7.72$$

And using the simplified equation, we get:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{SS_1+SS_2}{n^2-n}}} = \frac{51.6 - 90.6}{\sqrt{\frac{357.2+153.2}{5^2-5}}} = \frac{-39}{\sqrt{\frac{510.4}{20}}} = \frac{-39}{\sqrt{25.52}} = -7.72$$

So it works.

Also,

$$df = n_1 + n_2 - 2 = 10 - 2 = 8$$

In this example, we have 8 degrees of freedom, which gives us a critical t score of 2.306 for a two-tailed t-test at the .05 probability level. The absolute value of our calculated t score is 7.72, meaning that the differences between these two groups is significant at the .05 probability level. Furthermore, looking at our critical t score table, we can see that these differences are even significant at the .001 probability level, meaning that there is less than a 0.1% chance that these differences in scores are simply due to error or chance. Here, we could say the following: The difference in scores between our two groups was statistically significant ($t = -7.72$, $df = 8$, $p < .001$).

To calculate a t score for a dependent or within-subjects t-test, we need to use the following equation:

$$t = \sqrt{\frac{n - 1}{\left(\frac{n \sum D^2}{\left(\sum D\right)^2}\right) - 1}}$$

Here,

n = sample size

D = difference in scores for the respondent between Time 1 and Time 2, or between the matched pair

Say we had a class of five students, and they took the SAT (Scholastic Aptitude Test) before and after an extensive training course, and these were their scores:

Case	Score at Time 1	Score at Time 2
1	1250	1375
2	1170	1450
3	890	1250
4	1350	1495
5	750	1220

First, we would need to calculate the difference and the difference squared for each pair of scores:

Case	Score at Time 1	Score at Time 2	Difference	Difference Squared
1	1250	1375	–125	15625
2	1170	1450	–280	78400
3	890	1250	–360	129600
4	1350	1495	–145	21025
5	750	1220	–470	220900
Sum	—	—	–1380	465550

Plugging these values into our equation, we get the following:

$$t = \sqrt{\frac{n-1}{\left(\frac{n\sum D^2}{\left(\sum D\right)^2}\right)-1}} = \sqrt{\frac{5-1}{\left(\frac{5\times 465550}{1380^2}\right)-1}} = \sqrt{\frac{4}{\left(\frac{2327750}{1904400}\right)-1}}$$

$$= \sqrt{\frac{4}{0.22}} = \sqrt{17.99} = 4.24$$

For dependent samples *t*-tests,

$$df = n - 1$$

where n = the number of matched cases or pairs.

So for this example, we have 4 degrees of freedom. Using the .05 probability level, our critical t score is 2.776 for a two-tailed *t*-test. Since our calculated t score of 4.24 is greater than the critical t score of 2.776, the differences in scores from Time 1 to Time 2 are significant at the .05 probability level (i.e., the increase in scores was statistically significant at the .05 level). However, our calculated t score is not greater than the critical t value at the .01 probability level, 4.604. Here, we could say the following: The increase in SAT scores for our sample of five individuals from Time 1 to Time 2 was statistically significant ($t = 4.24$, $df = 4$, $p < .05$).

ANOVA: Theory

The ANOVA, which stands for analysis of variance, is like a generalized version of the *t*-test that can be used to test the difference in a continuous dependent variable between three or more groups or to test the level of a continuous dependent variable in a single group of respondents who were tested at three or more points in time. The *t*-test was published by William Sealy Gosset in 1908 under the pen name *Student*, which is why the *t*-test is sometimes referred to as the Student's *t*-test. The ANOVA was developed several decades later by Sir Ronald Fisher, which is why the ANOVA is sometimes called Fisher's ANOVA.

While the *t*-test relies on the *t* statistic, the ANOVA uses what is called the *F* statistic or *F*-test. When comparing two groups, either the *t*-test or the ANOVA may be used as they will both give you the same results. For example, below are the results from Stata for both the *t*-test and an ANOVA on years of education by gender for cases from the year 2004. You can see that the probability levels for both analyses are the same.

```
. ttest educ if year==2004, by(sex)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |    1279    13.81939    .0842611    3.013439    13.65408     13.9847
       2 |    1531      13.597    .0710005    2.778106    13.45773    13.73626
---------+--------------------------------------------------------------------
combined |    2810    13.69822    .0545035    2.889202    13.59135    13.80509
---------+--------------------------------------------------------------------
    diff |            .2223947    .1093871                .0079075    .4368819
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =   2.0331
Ho: diff = 0                                     degrees of freedom =     2808

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9789         Pr(|T| > |t|) = 0.0421          Pr(T > t) = 0.0211


. oneway educ sex if year==2004

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      34.4657999      1   34.4657999      4.13     0.0421
 Within groups      23413.6253   2808   8.33818565
------------------------------------------------------------------------
    Total           23448.0911   2809   8.34748704

Bartlett's test for equal variances:  chi2(1) =   9.2396  Prob>chi2 = 0.002
```

Calculating an ANOVA is slightly more complicated than calculating a *t*-test. Say we gave a survey to 10 whites, 10 blacks, and 10 Hispanics asking about their highest year of education completed, and we got the following data:

Whites	Blacks	Hispanics
14	12	14
12	14	16
16	12	10
20	12	10
12	12	14
12	10	12
16	8	12
16	10	12
14	12	8
20	20	8

First, we will calculate the following values for each group:

$\sum x$: a sum of all the scores of that group

$\bar{X}$: the mean of that group's scores

$\sum(x^2)$: a sum of the square of the group's scores

Next, we will calculate these values for the entire set of cases:

$\sum(\sum x)$: summing the three values for $\sum x$ that we computed previously

$\sum[\sum(x^2)]$: summing the three values for $\sum(x^2)$ that we computed previously

For example, we would get these values for whites:

$$\sum x = 14 + 12 + 16 + 20 + 12 + 12 + 16 + 16 + 14 + 20 = 152$$

$$\bar{x} = \frac{152}{10} = 15.2$$

$$\sum (x^2) = 14^2 + 12^2 + 16^2 + 20^2 + 12^2 + 12^2 + 16^2 + 16^2 + 14^2 + 20^2 = 2392$$

Doing the same computations for the other two groups would give you the following values:

Stat.	Whites	Blacks	Hispanics
$\sum x$	152.0	122.0	116.0
$\bar{x}$	15.2	12.2	11.6
$\sum(x^2)$	2392.0	1580.0	1408.0

Then,

$$\sum\left(\sum x\right) = 152 + 122 + 116 = 390$$

$$\sum\left(\sum\left(x^2\right)\right) = 2392 + 1580 + 1408 = 5380$$

Now, we need to calculate three different sum of squares values: the sum of squares total, the sum of squares between, and the sum of squares within. Then, we will compute the mean squares for between groups and within groups. Finally, we will compute the F statistic by dividing the mean squares between by the mean squares within.

So to begin,

$$\begin{aligned} SS\text{ total} &= \sum\left(\sum\left(x^2\right)\right) - \frac{\left[\sum\left(\sum x\right)\right]^2}{N} = 5380 - \frac{390^2}{30} \\ &= 5380 - \frac{152100}{30} = 5380 - 5070 = 310 \end{aligned}$$

$$\begin{aligned} SS\text{ between} &= \sum\frac{\left(\sum x\right)^2}{n} - \frac{\left[\sum\left(\sum x\right)\right]^2}{N} = \left(\frac{152^2}{10} + \frac{122^2}{10} + \frac{116^2}{10}\right) \\ &\quad - \left(\frac{390^2}{30}\right) = 5144.4 - 5070 = 74.4 \end{aligned}$$

$$SS\text{ within} = SS\text{ total} - SS\text{ between} = 310 - 74.4 = 235.6$$

As a check,

$$\begin{aligned} SS\text{ within} &= SS\text{ total for Group 1} + SS\text{ total for Group 2} \\ &\quad + SS\text{ total for Group 3} \\ &= \left(\sum\left(x^2\right) - \frac{\left(\sum x\right)^2}{N}\right) + \left(\sum\left(x^2\right) - \frac{\left(\sum x\right)^2}{N}\right) \\ &\quad + \left(\sum\left(x^2\right) - \frac{\left(\sum x\right)^2}{N}\right) \\ &= \left(2392 - \frac{152^2}{10}\right) + \left(1580 - \frac{122^2}{10}\right) + \left(1408 - \frac{116^2}{10}\right) \\ &= 81.6 + 91.6 + 62.4 = 235.6 \end{aligned}$$

Next,

$$MS\text{ between} = \frac{SS\text{ between}}{df\text{ between}} = \frac{74.4}{n(\text{groups}) - 1} = \frac{74.4}{3 - 1} = \frac{74.4}{2} = 37.2$$

$$MS\text{ within} = \frac{SS\text{ within}}{df\text{ within}} = \frac{235.6}{N - n(\text{groups})} = \frac{235.6}{30-3} = \frac{235.6}{27} = 8.726$$

Finally,

$$F = \frac{MS\text{ between}}{MS\text{ within}} = \frac{37.2}{8.726} = 4.263$$

Done.

Now, we need to consult an F table containing critical F values to see whether our results are significant or not. In our example, we had 2 degrees of freedom in the numerator (MS between) and 27 degrees of freedom in the denominator (MS within). Looking at an F table, this would give us a critical F value of approximately 3.38 at the .05 probability level. As you can see, our results were significant at the .05 probability level as our calculated F value, 4.263, was greater than the critical F value for the .05 probability level, 3.38. Here, we could say the following: There is a significant difference in the level of education between whites, blacks, and Hispanics ($F(2, 27) = 4.26$, $p < .05$). The first value for our degrees of freedom, 2, is equal to the number of groups minus one. The second value, 27, is equal to the total sample size or number of respondents, 30, minus the number of groups, 3.

If you wanted to combine all these steps into one, you would get the following equation for the F statistic:

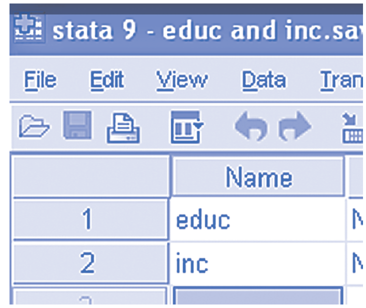
$$F = \frac{MS\text{ between}}{MS\text{ within}} = \frac{\frac{SS\text{ between}}{df\text{ between}}}{\frac{SS\text{ within}}{df\text{ within}}} = \frac{\frac{SS\text{ between}}{df\text{ between}}}{\frac{SS\text{ total} - SS\text{ between}}{df\text{ within}}}$$

$$= \frac{\left(\frac{\sum\frac{\left(\sum x\right)^2}{n} - \frac{\left[\sum\left(\sum x\right)\right]^2}{N}}{n(\text{groups}) - 1}\right)}{\left(\frac{\left[\sum\left(\sum\left(x^2\right)\right) - \frac{\left[\sum\left(\sum x\right)\right]^2}{N}\right] - \left[\sum\frac{\left(\sum x\right)^2}{n} - \frac{\left[\sum\left(\sum x\right)\right]^2}{N}\right]}{N - n(\text{groups})}\right)}$$

There are several versions of the ANOVA that will be covered in the SPSS and Stata sections of this chapter. The first, which was just presented, is called a **one-way ANOVA**. A one-way ANOVA is used when you have only one categorical independent or **predictor variable**. A **factorial ANOVA** is used when you have two or more categorical independent or predictor variables. Finally, a **repeated measures ANOVA** is used when you are looking at scores on a dependent variable across two or more points in time.

Pearson's *r*: IBM SPSS

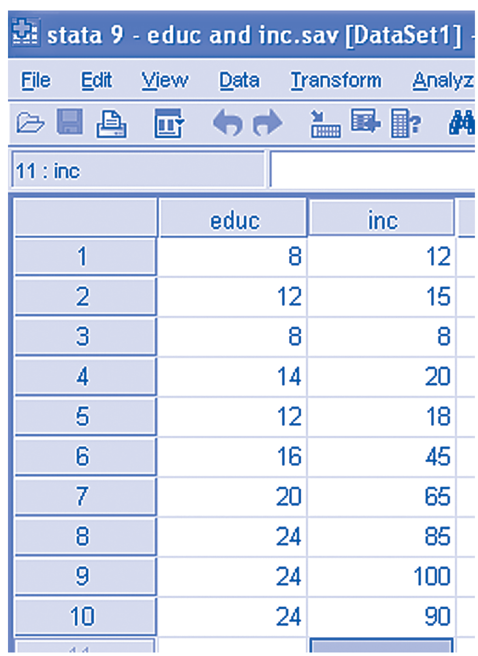
In this section, I will use the example from the previous section on Pearson's *r*, which determined the correlation coefficient between years of education and income. First, we will create two new variables in IBM SPSS, one called *educ* and another called *inc*, like the following:



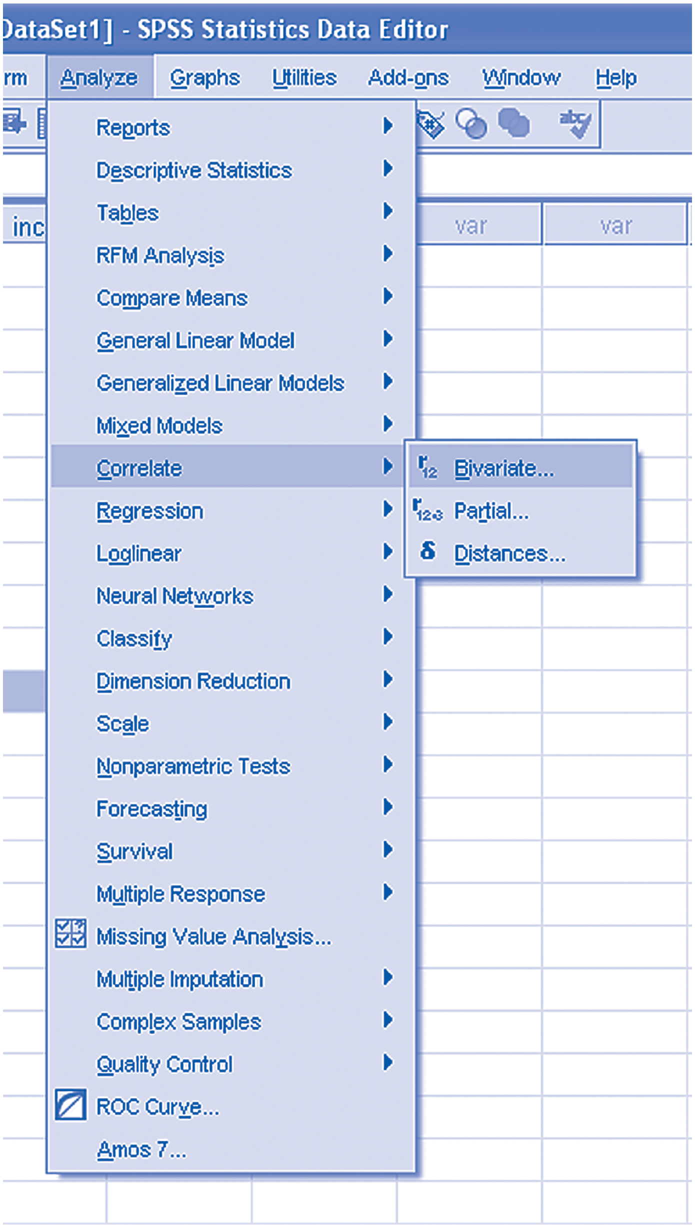
Next, we will enter the data, reproduced below, into IBM SPSS:

Years of Education	Income (in Thousands of $)
8	12
12	15
8	8
14	20
12	18
16	45
20	65
24	85
24	100
24	90

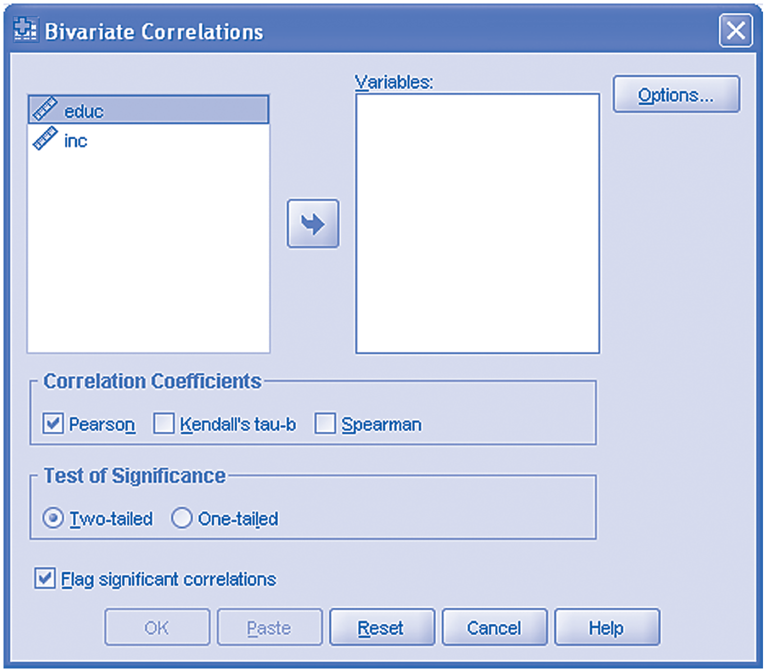
When you are finished, the *Data View* of IBM SPSS should look like this:



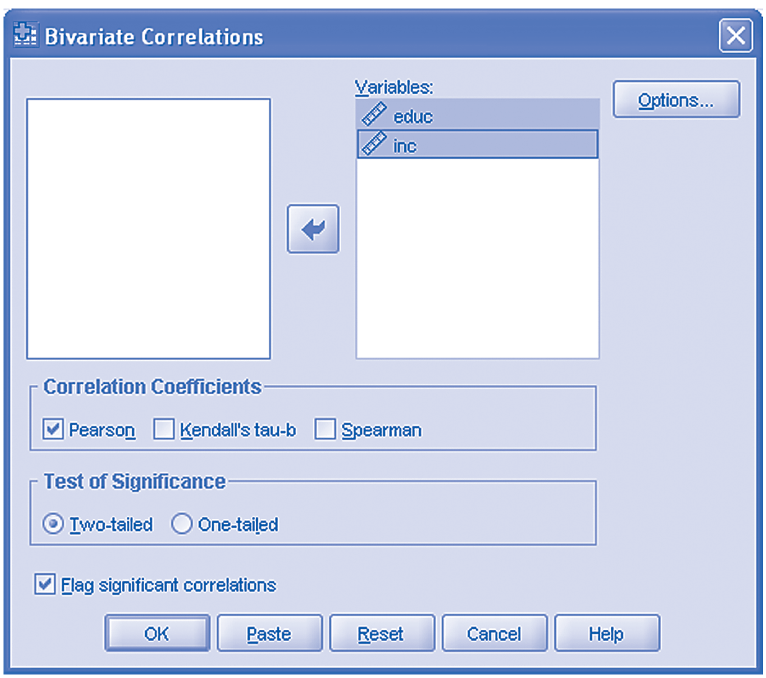
Next, make the following menu selection:



This will reveal the following dialog box:



Next, we will move our two variables over to the *Variables* box, like so:



We can leave all the other options as they are. Finally, clicking the *OK* button will give us the following results:

Correlations

```
[DataSet1] T:\Md book\#Data\Chapter 4\stata 9 - educ and inc.sav
```

Correlations

		educ	inc
educ	Pearson Correlation	1	.974**
	Sig. (2-tailed)		.000
	N	10	10
inc	Pearson Correlation	.974**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

As you can see, these results match what we determined by hand in the previous section. IBM SPSS has calculated the correlation coefficient between years of education and income to be 0.974 with a *p* level of less than .001 (as indicated by the ".000" under "Sig. (2-tailed)"). Here, we could say the following: There is a statistically significant positive correlation between years of education and income ($r = .97$, $p < .001$). As a note, whenever IBM SPSS gives your *p* level ("Sig.") as ".000," this means it is actually less than .001 but not equal to 0. For example, here, it might have been .0002 or .00005. SPSS basically just "rounds down" to zero in these cases.

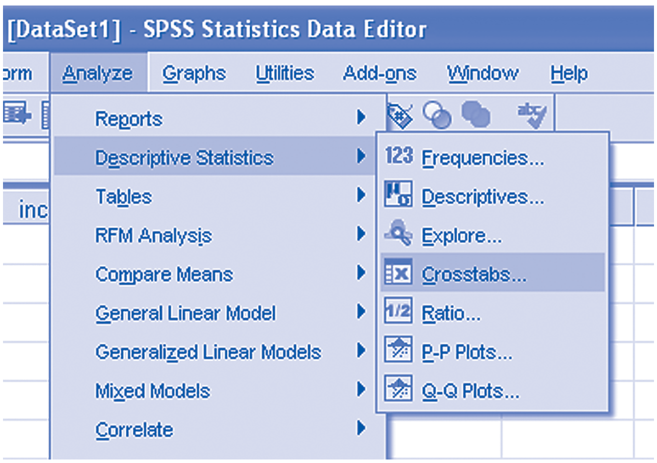
The corresponding syntax is presented below:

```
CORRELATIONS
/VARIABLES=educ inc
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

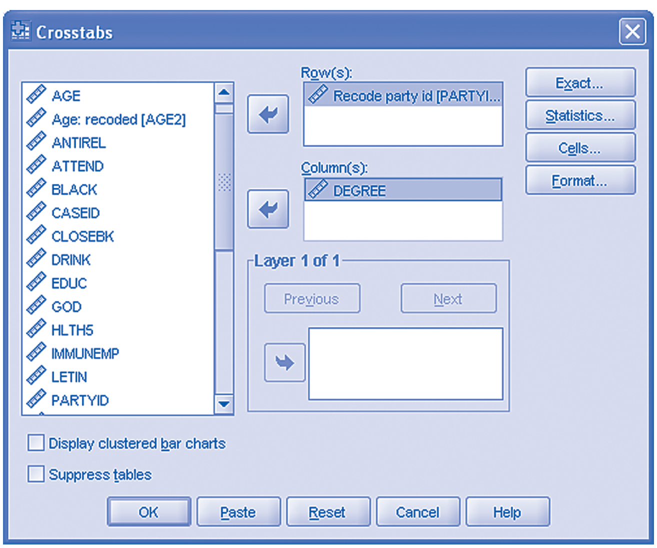
Chi-Square: IBM SPSS

Calculating the chi-square statistic in IBM SPSS is very quick and easy, and it is obviously preferred to calculating it by hand. In our examples here, we will look at the relationship between highest degree completed and political affiliation, using actual data from the General Social Survey (GSS).

First, navigate to the following menu selection:

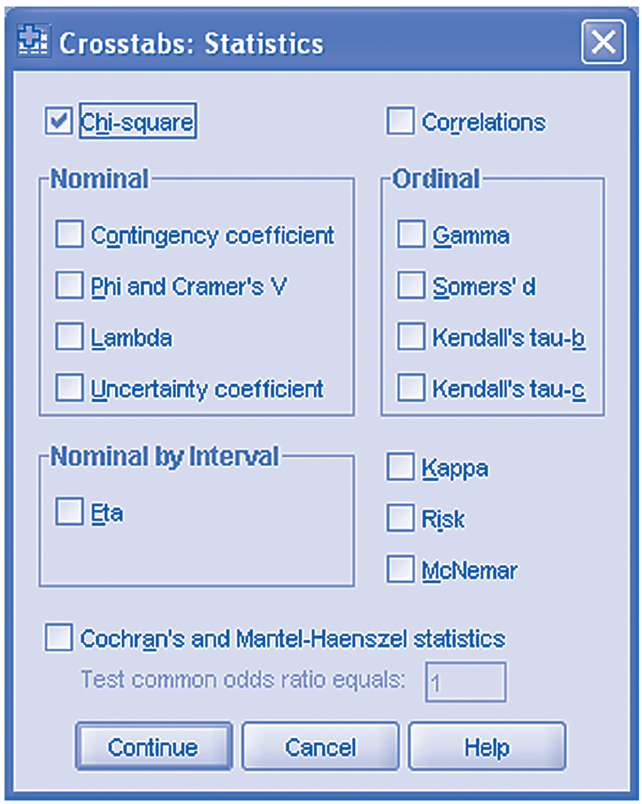


This will bring up the following dialog box:



Here, I have selected a recoded version of *partyid* (political party affiliation) under *Row*, and *degree* (highest degree completed) under *Column*. In case you are following along yourself, I recoded respondents who answered 0 through 2 (strong Democrat, not very strong Democrat, or independent close to Democrat) as "1," those who responded 3 (independent) as "2," and those responded 4 through 6 (independent close to Republican, not very strong Republican, strong Republican) as "3." Those who responded 7 (other party or refused), or were missing (8 or 9), I recoded as missing.

Next, you will want to click on *Statistics*. This will bring up the following dialog box:



All you need to do here is to check the *Chi-square* box. Clicking *Continue* and *OK* will result in the following output:

Crosstabs

```
[DataSet2] T:\Md book\#Data\Chapter 3 GSS data\data.sav
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Recode party id * DEGREE	45462	97.7%	1048	2.3%	46510	100.0%

Recode party id * DEGREE Crosstabulation

Count

		DEGREE					Total
		Less than high school	High school	Associate/Junior College	Bachelor's	Graduate	
Recode party id	Democrat	6334	11563	1045	2618	1431	22991
	Independent	1698	3529	303	601	286	6417
	Republican	2832	8513	825	2812	1072	16054
Total		10864	23605	2173	6031	2789	45462

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	834.007[a]	8	.000
Likelihood Ratio	847.994	8	.000
Linear-by-Linear Association	388.345	1	.000
N of Valid Cases	45462		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 306.72.

The calculated chi-square value in this example was 834.007. We had 8 degrees of freedom ((3 – 1) × (5 – 1)). IBM SPSS tells us that this was significant at the .001 probability level. Here, we could say the following: There was a statistically significant relationship between highest degree completed and political party affiliation ($\chi^2 = 834.0$, $df = 8$, $p < .001$).

This is the corresponding syntax for this example:

```
DATASET ACTIVATE DataSet2.
CROSSTABS
/TABLES=PARTYIDR BY DEGREE
```

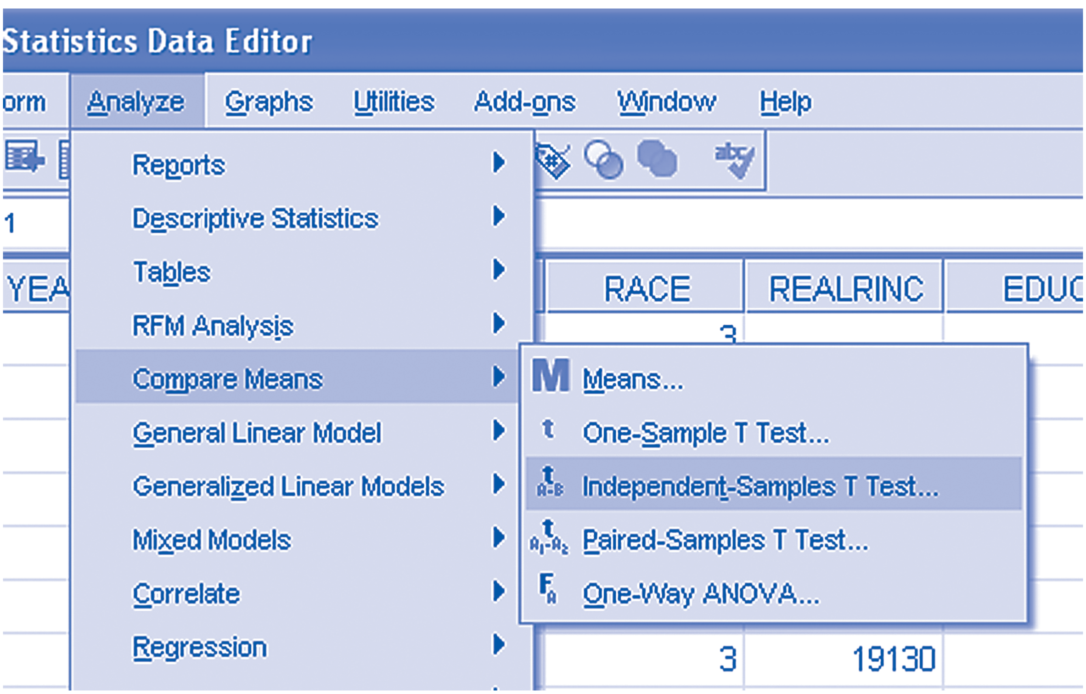
```
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ
/CELLS=COUNT
/COUNT ROUND CELL.
```

And if you wanted to omit the crosstabulation table, you would use this syntax:

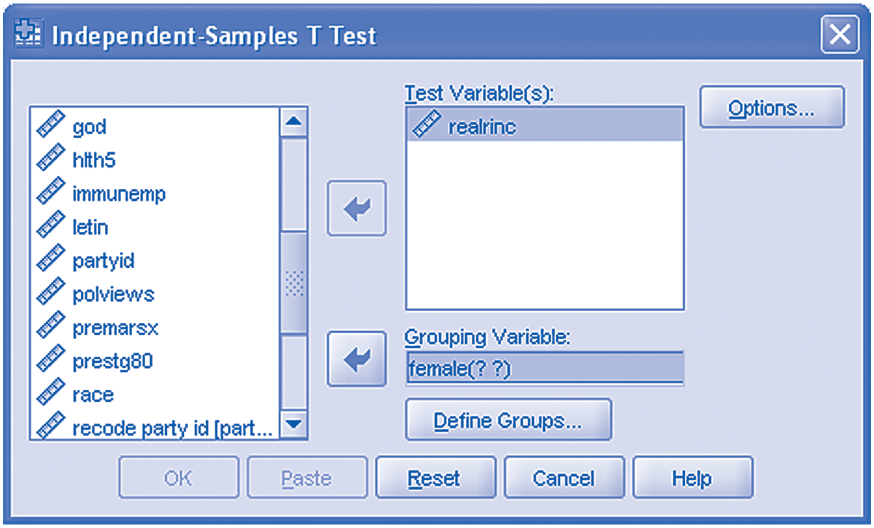
```
CROSSTABS
/TABLES=PARTYIDR BY DEGREE
/FORMAT=NOTABLES
/STATISTICS=CHISQ
/COUNT ROUND CELL.
```

t-Test: IBM SPSS

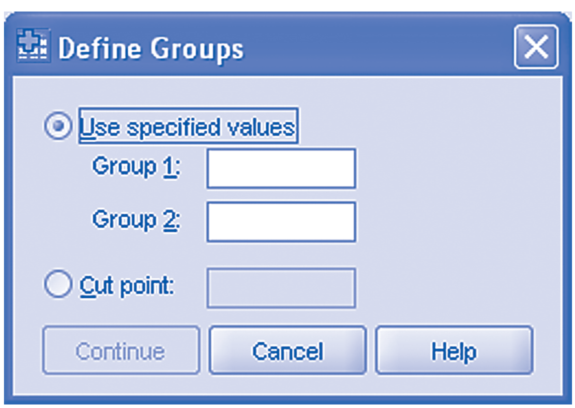
To run an independent samples *t*-test within IBM SPSS, we will choose the following menu selection:



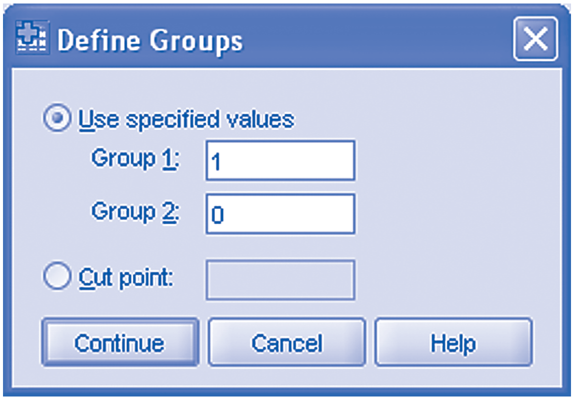
This will open the following dialog box:



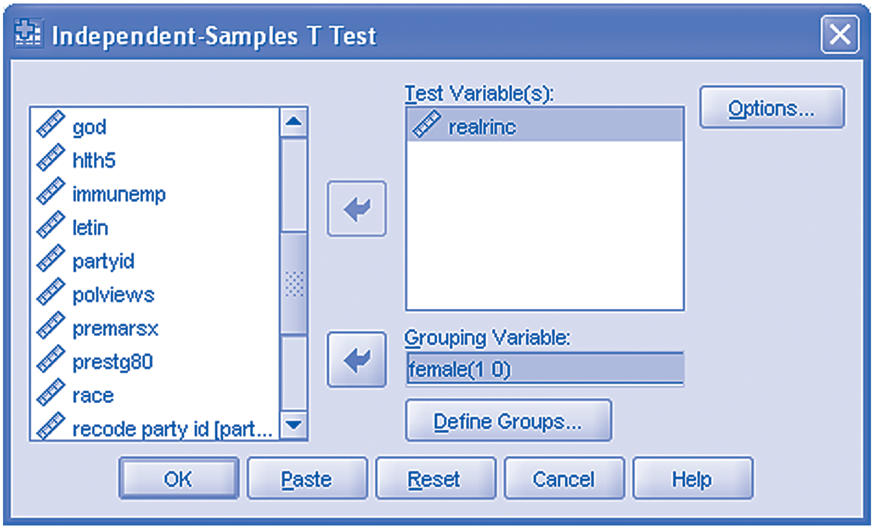
As you can see, I have taken the liberty of adding the respondents' yearly income, *realrinc*, into the *Test Variable(s)* box, and adding *female* (a constructed **dummy variable**, where *female* = 1 and *male* = 0) into the *Grouping Variable* box. Right now, inside the parentheses next to the sex variable, there are two question marks. Before we can run the *t*-test within IBM SPSS, we click on the *Define Groups* button. Clicking on the button will reveal this dialog box:



Here, we specify the values for the two different groups (males and females). Since females are coded 1 and males are coded 0, I simply specified Group 1 as equal to 1 (*females*), and Group 2 as equal to 0 (*males*), like this:



Now, you can see that our two groups are defined correctly:



Clicking *OK* will result in the following output:

T-Test

```
[DataSet1] T:\Md book\#Data\Chapter 3 GSS data\gss data.sav
```

Group Statistics

	female	N	Mean	Std. Deviation	Std. Error Mean
realrinc	1	13301	14811.22	14022.518	121.586
	0	13862	26983.16	24339.702	206.729

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
realrinc	Equal variances assumed	1318.087	.000	-50.228	27161	.000	-12171.943	242.331	-12646.925	-11696.962
	Equal variances not assumed			-50.752	22324.904	.000	-12171.943	239.834	-12642.034	-11701.852

First, we see that Levene's test for equality of variances was significant at the .05 probability level. This means that the variances between groups are significantly different, and therefore when looking up the *t* values and significance, we should use the second row labeled "Equal variances not assumed." Here, we see that we obtained a *t* score of –50.752 with 22324.9 degrees of freedom, which was significant at the .001 probability level. In IBM SPSS, if the probability level or level of significance is ever listed as ".000," this means that it is less than .001—this is an IBM SPSS bug. The results also show us the mean for the two different groups: the mean income for males is approximately $26,983 per year, while the mean income for females is approximately $14,811 per year. These results could be stated as follows: Males were found to have a significantly higher income as compared with female respondents ($t = -50.75$, $df = 22324.90$, $p < .001$). Keep in mind that this analysis includes all data, starting in the year 1972. If we include only cases from the year 2004, we get the following results:

T-Test

```
[DataSet1] T:\Md book\#Data\Chapter 3 GSS data\gss data.sav
```

Group Statistics

	female	N	Mean	Std. Deviation	Std. Error Mean
realrinc	1	833	19191.23	21500.991	744.965
	0	855	31380.56	32430.227	1109.090

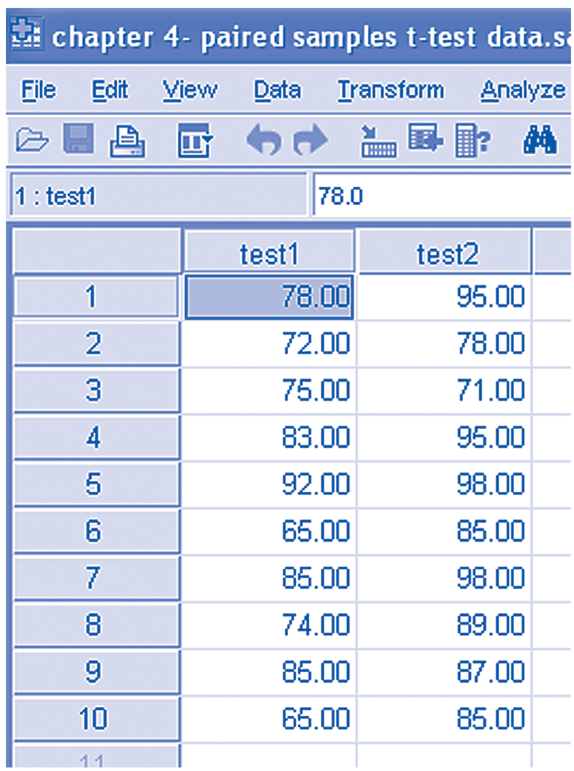
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
realrinc	Equal variances assumed	68.334	.000	-9.077	1686	.000	-12189.334	1342.860	-14823.182	-9555.487
	Equal variances not assumed			-9.123	1487.614	.000	-12189.334	1336.059	-14810.095	-9568.574

And this is the corresponding syntax:

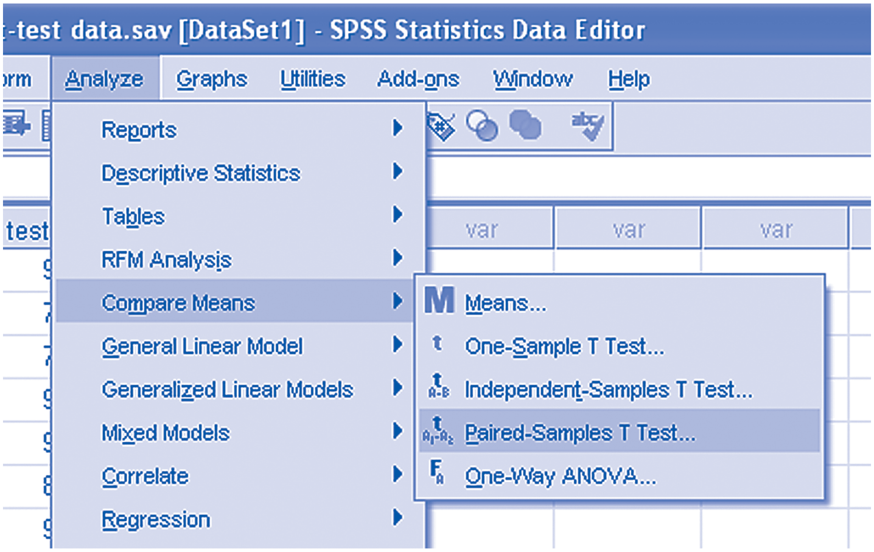
```
T-TEST GROUPS=female(1 0)
/MISSING=ANALYSIS
/VARIABLES=realrinc
/CRITERIA=CI(.95).
```

Now, let's use IBM SPSS to run a dependent samples *t*-test. Because the GSS does not contain any variables that would be appropriate to use in a dependent samples *t*-test, I simply created a new file within IBM SPSS and created two new variables: *test1* and *test2*. Then, I typed in the following data as an example:

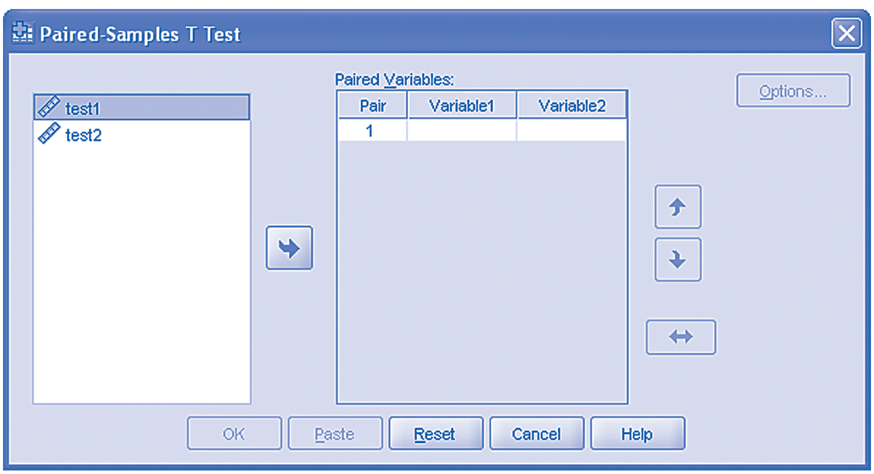


	test1	test2
1	78.00	95.00
2	72.00	78.00
3	75.00	71.00
4	83.00	95.00
5	92.00	98.00
6	65.00	85.00
7	85.00	98.00
8	74.00	89.00
9	85.00	87.00
10	65.00	85.00

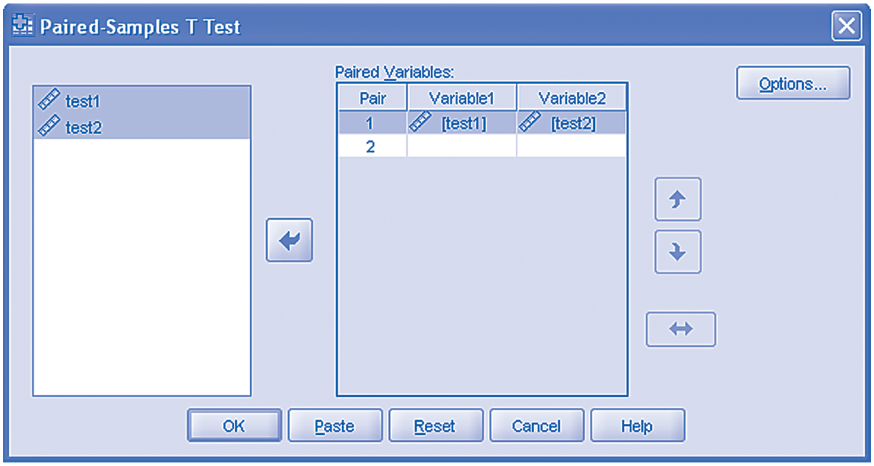
Simply input the same data values if you want to follow along in IBM SPSS. Next, you will navigate to the following menu selection:



Next, the following dialog box will appear:



In this dialog box, I've simply selected the variables *test1* and *test2* and moved them to the *Paired Variables* box on the right, like this:



After clicking *OK*, we get the following output:

T-Test

```
[DataSet1] T:\Md book\#Data\Chapter 4\chapter 4- paired samples t-test data.sav
```

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	test1	77.4000	10	8.90942	2.81741
	test2	88.1000	10	8.86253	2.80258

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	test1 & test2	10	.596	.069

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	test1 - test2	-10.70000	7.98679	2.52565	-16.41341	-4.98659	-4.237	9	.002

IBM SPSS calculated the t score in this example to be –4.237. With 9 degrees of freedom, our results are significant at the .002 probability level (2-tailed). When writing up these results, you would simply say that it was significant at the $p < .01$ level: you will only use .05, .01, or .001 as standards. For example, you could report this result in the following way: Scores on Test 2 were found to be significantly higher as compared with scores on Test 1 ($t = -4.24$, $df = 9$, $p < .01$).

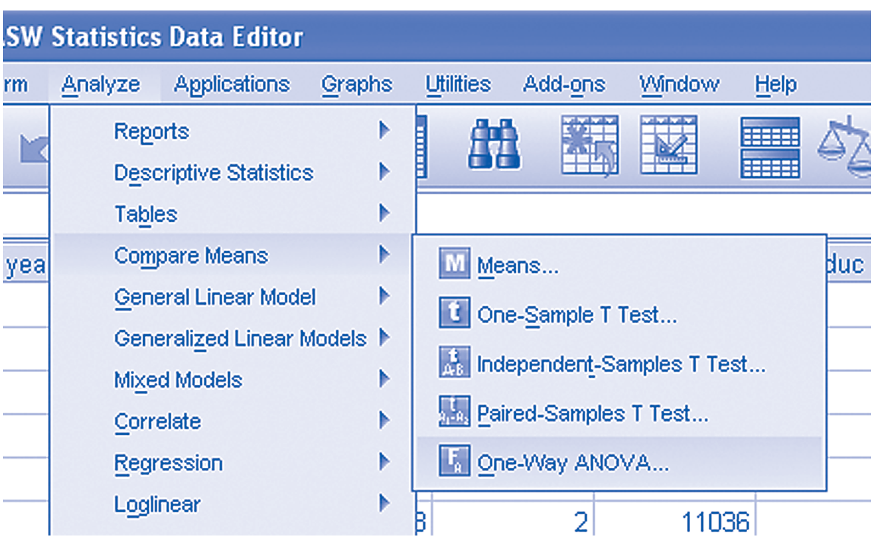
This is the corresponding syntax:

```
T-TEST PAIRS=test1 WITH test2 (PAIRED)
/CRITERIA=CI(.9500)
/MISSING=ANALYSIS.
```

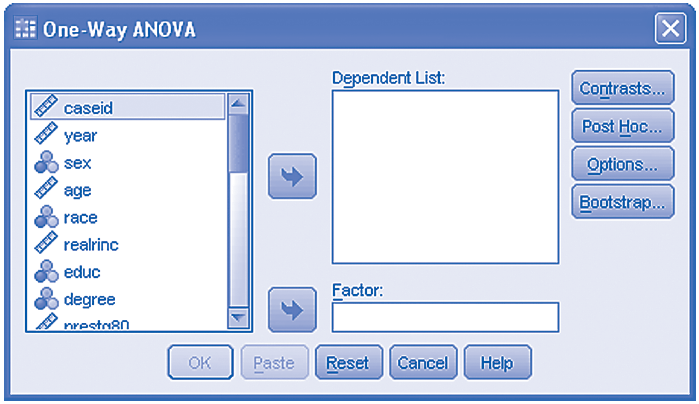
One-Way ANOVA: IBM SPSS

Let's try running an ANOVA on differences in years of education based on race, similar to the race example that was presented in the theoretical section. In this example, race is coded as three categories: blacks, whites, and those of other race. These data are derived from the GSS, a large national survey of American adults.

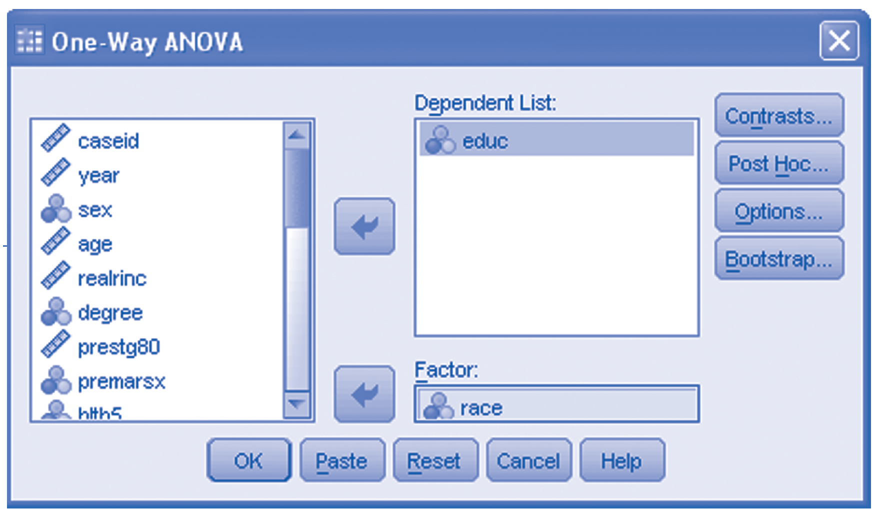
First, navigate to the following menu selection:



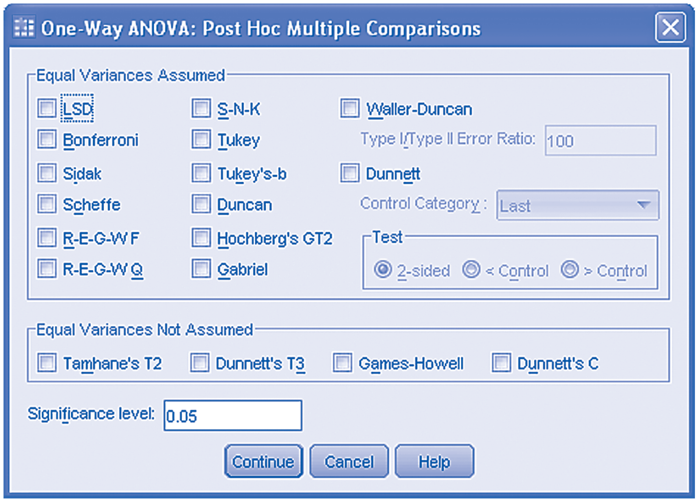
This will open the following dialog box:



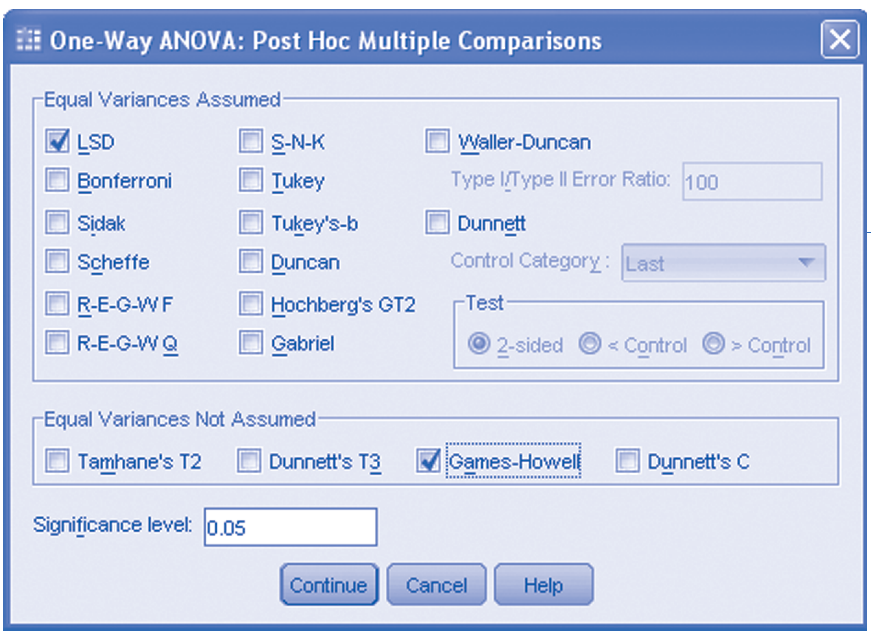
Here, I will simply add the dependent variable *educ*, representing the highest year of education completed, in the *Dependent List* and *race* into the *Factor* box, like this:

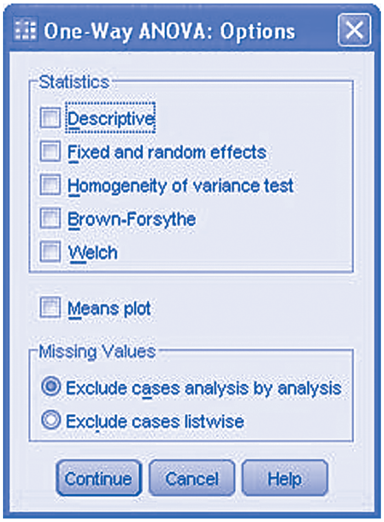


Next, you'll want to click on the *Post Hoc* button. This dialog box will pop up:



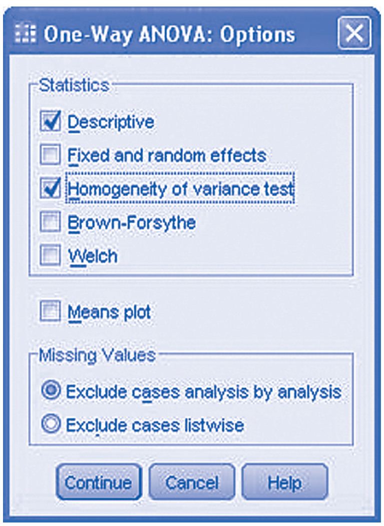
In this window, I will simply select two post hoc tests, the LSD (least significant difference) and the Games-Howell post hoc tests:





If you do not run a post hoc test, you will not know between which specific groups there is a statistically significant difference. For example, even if the *F* test for the ANOVA is significant, we will not know whether all three groups differ from each other significantly in their scores, if there is only a significant difference between whites and blacks, and so on. To ascertain between which specific groups there is a significant difference, we need to run a post hoc analysis in addition to the ANOVA. If you do additional reading into the different post hoc analyses that are available when conducting an ANOVA, you will find that they differ in particular ways, especially in terms of how conservative they are. Some are also more appropriate for particular types of situations: for example, when your ANOVA includes a small number of groups or a large number of groups.

The LSD post hoc test is less conservative, while the Games-Howell post hoc test is more conservative. As you can see, SPSS includes two categories, a large set of tests under "Equal Variances Assumed," and a smaller set under "Equal Variances Not Assumed." In our example, if the variance in years of education is significantly different between whites, blacks, and members of other races, we should choose one of the four post hoc tests under "Equal Variances Not Assumed." We can test whether the variances are significantly different in the following way: first, click *Continue* to close out this dialog box. Next, click *Options*. This reveals the dialog box to the left.



Here, I will select *Homogeneity of variance test*, which will test whether the variance of level of education is significantly different across race. I have also selected *Descriptive*, which will give the mean years of education for whites, blacks, and members of other races separately.

After clicking *Continue* and *OK*, you will see the following results:

Descriptives

educ

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
White	38383	12.76	3.095	.016	12.72	12.79	0	20
Black	6360	11.64	3.311	.042	11.56	11.72	0	20
Other	1626	12.90	3.562	.088	12.73	13.08	0	20
Total	46369	12.61	3.167	.015	12.58	12.64	0	20

Test of Homogeneity of Variances

educ

Levene Statistic	df1	df2	Sig.
23.817	2	46366	.000

ANOVA

educ

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6898.374	2	3449.187	349.095	.000
Within Groups	458112.554	46366	9.880		
Total	465010.928	46368			

Post Hoc Tests

Multiple Comparisons

Dependent Variable:educ

	(I) race	(J) race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LSD	White	Black	1.112*	.043	.000	1.03	1.20
		Other	-.147	.080	.064	-.30	.01
	Black	White	-1.112*	.043	.000	-1.20	-1.03
		Other	-1.260*	.087	.000	-1.43	-1.09
	Other	White	.147	.080	.064	.00	.30
		Black	1.260*	.087	.000	1.09	1.43
Games-Howell	White	Black	1.112*	.044	.000	1.01	1.22
		Other	-.147	.090	.229	-.36	.06
	Black	White	-1.112*	.044	.000	-1.22	-1.01
		Other	-1.260*	.098	.000	-1.49	-1.03
	Other	White	.147	.090	.229	-.06	.36
		Black	1.260*	.098	.000	1.03	1.49

*. The mean difference is significant at the 0.05 level.

In the first table, labeled "Descriptives," we see the mean of years of education for whites, blacks, members of other races, and all groups combined. While the differences are not huge, it does appear that whites and members of other races tend to have more education as compared with blacks.

The second table, labeled "Test of Homogeneity of Variances," reports Levene's test for the equality of variances. Our probability level, which is circled, was found to be less than .05, which means that the variances in the level of education are significantly different across race. This also means that we will select as a post hoc test an option that does not assume equal variances. In this example, I selected the Games-Howell post hoc test, which does not assume equal variances.

Before moving to the results of the post hoc test, let's first discuss the results of the ANOVA itself. We see that the *F* statistic was calculated by IBM SPSS to be 349.095, with 2 degrees of freedom between groups and 46366 degrees of freedom within groups. This was significant at the $p < .001$ level. As the *F* test in this ANOVA was found to be significant, this means that level of education differs significantly based on race. However, to ascertain between which groups specifically there is a significant difference in education, we need to look at the results of our post hoc test. In regard to the degrees of freedom, which will be reported when writing up the results of an ANOVA, the between-groups degrees of freedom, calculated here to be 2, is simply the total number of groups minus one. In this example, we had three categories of race, so the between-groups degrees of freedom is simply 3 minus 1. The within-groups degrees of freedom is calculated as the total sample size minus the number of groups. The total sample size for this ANOVA, reported in the final row of the "Descriptives" table under *N*, was 46369. As we had three groups, the within-groups degrees of freedom is simply 46369 minus 3.

Finally, let's look at the results of our post hoc analysis, which are displayed under the "Post Hoc Tests" table. As you may notice, the results of our two post hoc analyses are quite similar, despite the fact that the LSD test assumes the equality of variances, while the Games-Howell test does not. This is not rare, as different tests commonly result in similar or identical results.

However, let's focus on the results of the Games-Howell test, as we found that the variances of level of education based on race significantly vary. Here, two results that were significant at the .05 probability level, denoted by asterisks, were found. First, whites were found to have significantly higher levels of education as compared with blacks. Specifically, whites were found, on average, to have 1.112 greater years of education as compared with blacks. Looking under the *Sig.* column, we can see that this was significant at the $p < .001$ level. As you may notice, the results of each comparison are actually reported twice in this table. Moving down two rows, the opposite comparison, blacks as compared with whites, is displayed. If you preferred, you could instead report this result, stating that blacks were found, on average, to have 1.112 fewer years of education as compared with whites. Just make sure to choose only one of these two results to report so you are not in effect reporting the same result twice.

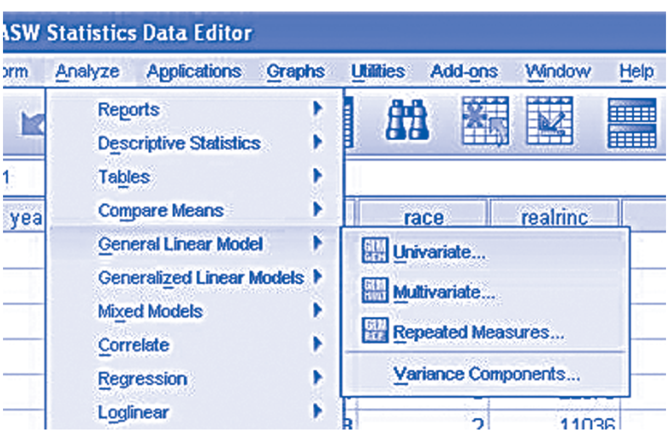
Finally, we can see in the final row of the table that members of other races were found, on average, to have 1.260 greater years of education as compared with blacks. Looking under the *Sig.* column, we can see that this was significant at the $p < .001$ level. Our results highlight the importance of running a post hoc test whenever we are conducting an ANOVA on more than two groups: While a significant difference was found between whites and blacks and between members of other races and blacks, no significant difference was found between whites and members of other races in regard to years of education.

Our results can be stated in the following way: A significant difference in years of education between whites, blacks, and members of other races was found, $F(2, 46366) = 349.10$, $p < .001$. Specifically, a Games-Howell post hoc test found the mean level of education for both whites and members of other races to be significantly greater than that of blacks, $p < .001$.

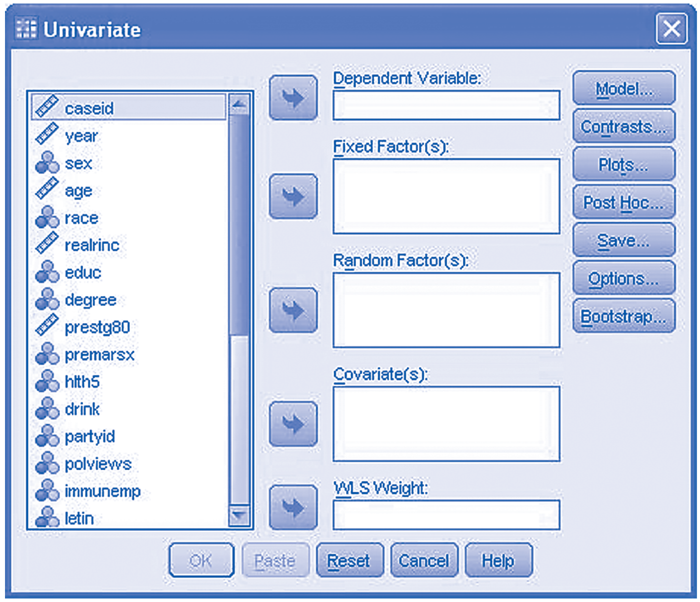
Factorial ANOVA: IBM SPSS

While one-way ANOVAs only include one categorical independent or predictor variable, factorial ANOVAs include more than one. In the example presented in the previous section, a one-way ANOVA was used as there was only one independent or predictor variable, race of the respondent. In this example, I will incorporate both race of the respondent as well as the respondent's gender in a factorial ANOVA that includes the respondent's income as the dependent variable. This example will also use data from the GSS.

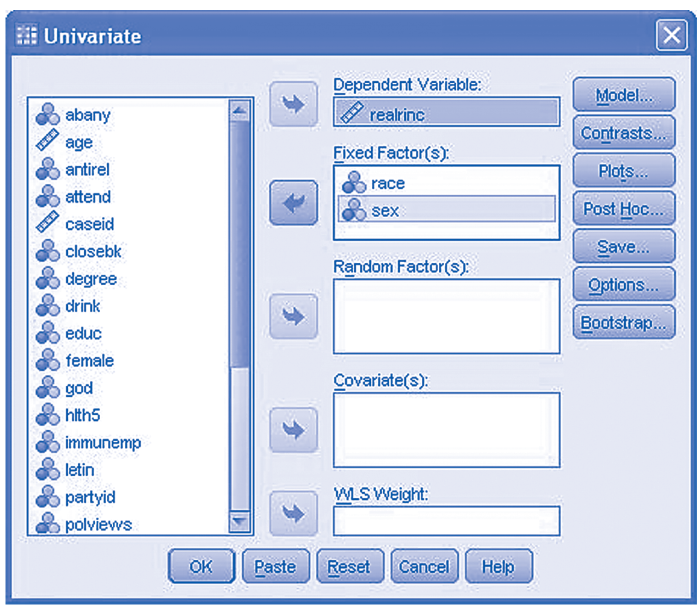
To begin, first make the following menu selection:



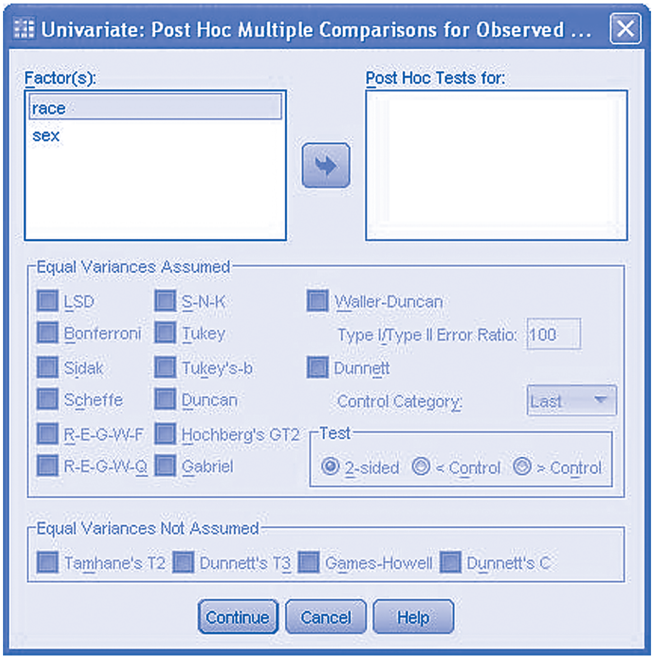
Next, the following dialog box will appear:



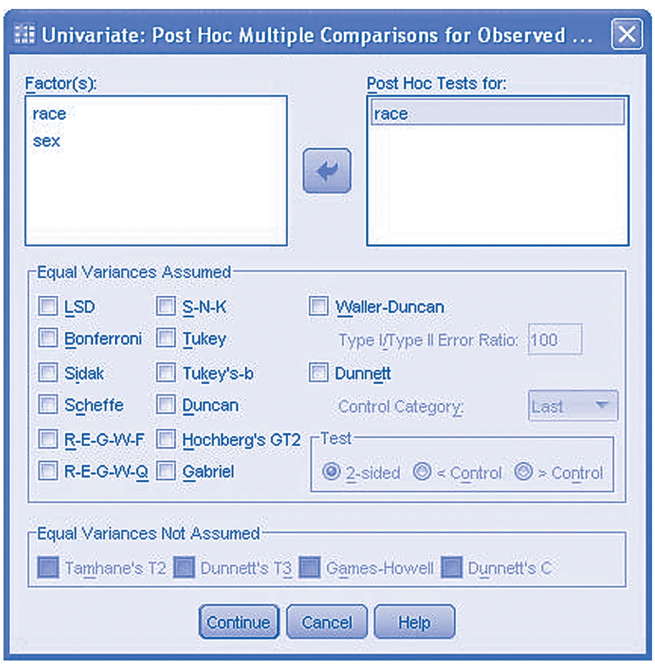
While this selection allows us to run a number of different tests, in this example, it will be used to run a factorial ANOVA. Next, I will move income, *realrinc*, into the *Dependent Variable* box and will move race and sex into the *Fixed Factor(s)* box, which is used for categorical independent variables. When completed, the dialog box will look as shown below.



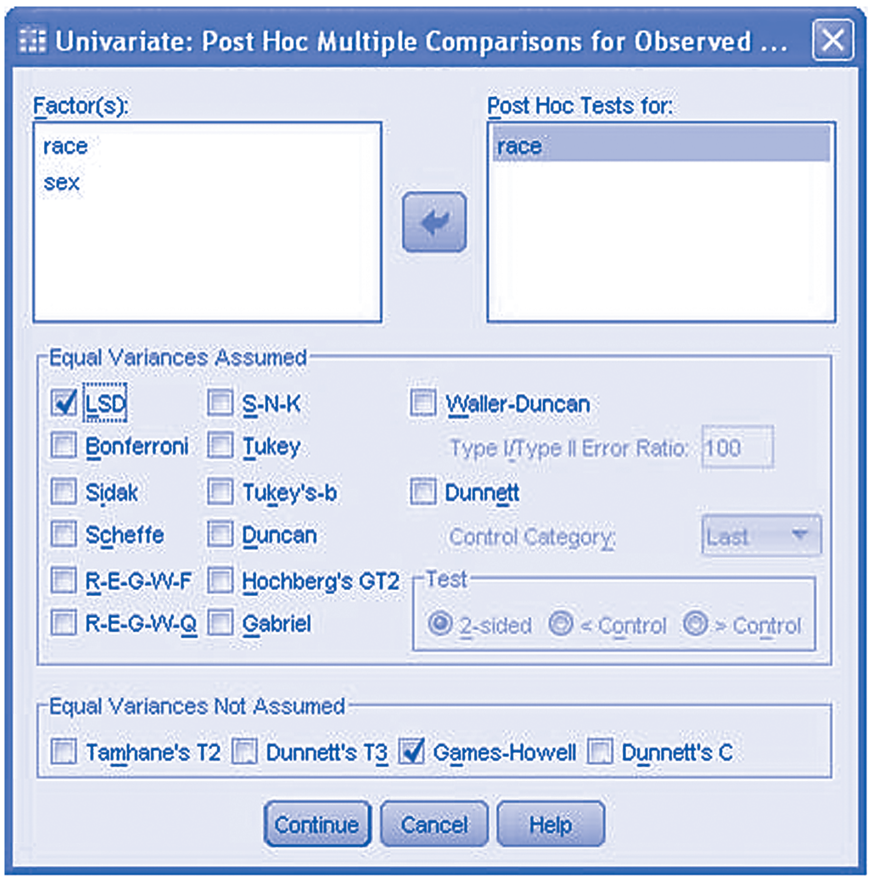
Next, click *Post Hoc*. This will open the following dialog box:



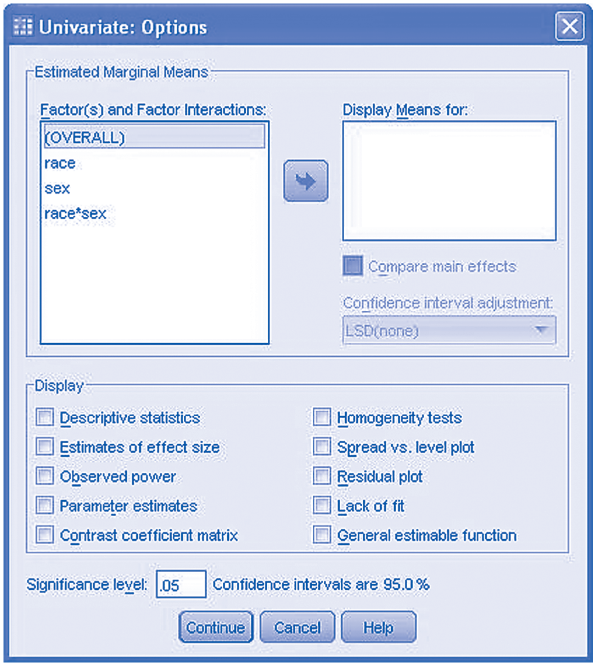
Here, I will move one of our independent variables, *race*, into the "Post Hoc Tests for" box:



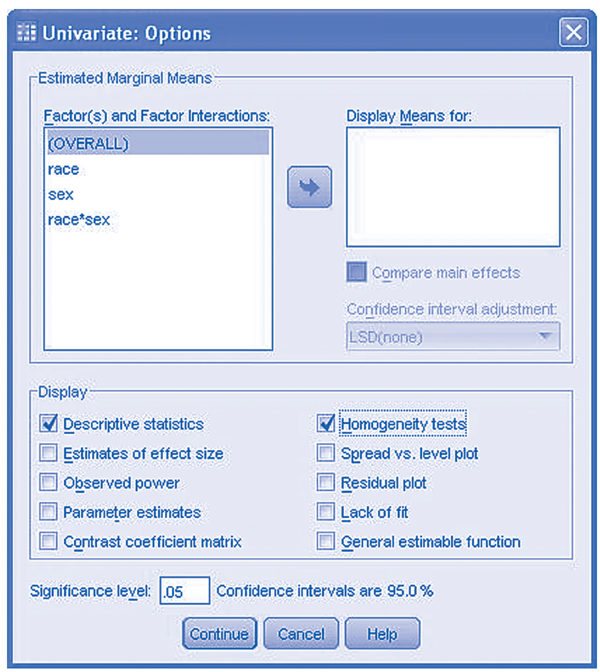
As there are only two categories for sex, male and female, a post hoc test is not necessary. As in the previous example, I will select both the LSD as well as the Games-Howell post hoc tests. The LSD post hoc test is less conservative, while the Games-Howell post hoc test is more conservative. More conservative tests are sometimes preferred, as you are less likely to get a "false positive," or a significant result, in situations where there actually is no real difference. As explained in the previous section, most post hoc tests assume that the variances in the dependent variable are not significantly different across categories of the independent variables. As we do not yet know whether this is the case, I will select one test from each category. After making these selections, our dialog box will look like the following:



Next, click *Continue*. Then, click *Options* to reveal the following dialog box:



Here, we will simply select "Homogeneity tests" to test whether the variance in income is significantly different across race and sex. This is important in determining which post hoc test we will use and report in this analysis. I'll also select "Descriptive statistics," which will give us the mean score on respondent's income by race and sex. After making the selections, our dialog box will look like the following:



Finally, click *Continue* and *OK* to run the analysis.

As the results of this analysis are rather lengthy, I'll go through them step by step as opposed to presenting the entire set of results all at once. The first two tables are presented on the following page.

Between-Subjects Factors

		Value Label	N
race	1	White	22507
	2	Black	3563
	3	Other	1093
sex	1	Male	13862
	2	Female	13301

Descriptive Statistics

Dependent Variable:realrinc

race	sex	Mean	Std. Deviation	N
White	Male	28176.87	25095.351	11766
	Female	15094.84	14427.765	10741
	Total	21933.74	21708.267	22507
Black	Male	19068.24	15621.340	1527
	Female	13284.99	11596.369	2036
	Total	15763.53	13768.218	3563
Other	Male	23540.04	23387.082	569
	Female	14927.64	13846.850	524
	Total	19411.13	19870.964	1093
Total	Male	26983.16	24339.702	13862
	Female	14811.22	14022.518	13301
	Total	21022.88	20871.594	27163

The first table, titled "Between-Subjects Factors," gives us the sample size, or number of respondents, for each category of our independent variables. For example, in the first row, we see that there are 22,507 white respondents included in this analysis. In the final row of this table, we see that there are 13,301 female respondents included in the analysis.

The second table, titled "Descriptive Statistics," gives us the mean for every possible combination of our independent variables. For example, in the first row, we see that the mean income for white males included in this analysis is $28176.87. If we wanted to find the mean income for females of other races, we simply find the *Other* category for race, which is the third one down, and then find female, which is the second row. Here, we see that the mean income for females of other races is $14927.64.

Next, I will present the tables for Levene's test of the equality of variances and the table presenting the main results of the ANOVA.

Levene's Test of Equality of Error Variances[a]

Dependent Variable:realrinc

F	df1	df2	Sig.
303.008	5	27157	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + race + sex + race * sex

Tests of Between-Subjects Effects

Dependent Variable:realrinc

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.130E12	5	2.261E11	573.719	.000
Intercept	2.609E12	1	2.609E12	6619.628	.000
race	9.294E10	2	4.647E10	117.926	.000
sex	1.513E11	1	1.513E11	383.954	.000
race * sex	4.342E10	2	2.171E10	55.086	.000
Error	1.070E13	27157	3.941E8		
Total	2.384E13	27163			
Corrected Total	1.183E13	27162			

a. R Squared = .096 (Adjusted R Squared = .095)

The first table here gives us the results of Levene's test of the equality of variances. This result was found to be significant at the $p < .001$ level, which means that the variance in income significantly varies across the categories of our independent variables and also means that we will select a post hoc test that does not assume equal variances.

The second table, titled "Tests of Between-Subjects Effects," presents the main results of the ANOVA. The first row, titled *Corrected Model*, gives us the results of the F test for the overall model. Here, the calculated F statistic was 573.719 and was significant at the $p < .001$ level. The three other results that are circled in this table give us the effects of race on income, sex on income, and the interaction between race and sex on income. First, the F statistic for race was 117.926. This was significant at the $p < .001$ level, which means that respondent's income was found to significantly vary based on race. Next, the F statistic for sex was 383.954. This result was also significant at the $p < .001$ level, meaning that respondent's income significantly varies based on sex. Finally, the interaction between race and sex, denoted as *race * sex*, had a calculated F statistic of 55.086 and was also significant at the $p < .001$ level. This means that the effect of race on income significantly varies by sex. Alternatively, you could state that the effect of sex on income varies significantly by race. For example, this would be the case if race were an important predictor of income for males but not for females. Likewise, this would be the case if males have higher incomes than females for whites but if females had higher incomes than males for blacks. In essence, the significant interaction effect in this example means that the effect of one of the independent variables on the dependent variable varies significantly depending on the level of the second independent variable. Interaction effects can clearly be trickier to deal with and can take some additional time to fully understand. The degrees of freedom, which you will report, come from the *df* column in the table just presented. For example, the *F* test for the full model would be reported as the following: $F(5, 27157) = 573.72$. The first value, 5, comes from the first row, while the second value, 27157, comes from the *Error* row. As you can see in the results write-up presented at the end of this section, this second value will always be equal to the value presented in the *Error* row.

Finally, I'll present the table which included the results of the post hoc tests we conducted.

Post Hoc Tests

race

Multiple Comparisons

Dependent Variable:realrinc

	(I) race	(J) race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	White	Black	6170.21*	357.927	.000	5468.66	6871.77
		Other	2522.61*	614.862	.000	1317.45	3727.77
	Black	White	-6170.21*	357.927	.000	-6871.77	-5468.66
		Other	-3647.60*	686.403	.000	-4992.98	-2302.21
	Other	White	-2522.61*	614.862	.000	-3727.77	-1317.45
		Black	3647.60*	686.403	.000	2302.21	4992.98
Games-Howell	White	Black	6170.21*	272.289	.000	5531.91	6808.52
		Other	2522.61*	618.220	.000	1071.92	3973.31
	Black	White	-6170.21*	272.289	.000	-6808.52	-5531.91
		Other	-3647.60*	643.787	.000	-5158.02	-2137.18
	Other	White	-2522.61*	618.220	.000	-3973.31	-1071.92
		Black	3647.60*	643.787	.000	2137.18	5158.02

Based on observed means.
The error term is Mean Square(Error) = 394077231.261.

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

As mentioned previously, a post hoc test for sex was not necessary as there are only two groups, males and females. The results of the ANOVA, presented previously, found that respondent's income varied significantly based on sex. Looking at the "Descriptive Statistics" table, presented previously, we see that the average income for males is $26983.16, while the average income for females is $14811.22. Using this information, we can state that the average income for males is significantly higher than that of females.

Now, to look at the results presented in this table. First, as the variance in income was found to significantly differ across categories of our independent variables, we will focus only on the second post hoc test presented in this table, the Games-Howell post hoc test, as it does not assume equal variances, while the LSD test does. As you may notice, the results for these two tests are similar. However, we should focus on and report the results from the Games-Howell test as it does not assume equal variances. In this post hoc test, three significant comparisons were found, which means that there are significant differences in income between all three of our racial categories. As mentioned in the previous section, all comparisons are made twice, so all results are repeated. For example, the white versus black comparison had a mean difference of 6170.21, while the black versus white comparison had a mean difference of –6170.21. In essence, this is the same result, simply flipped, so when looking at this table, we can simply focus on positive mean differences, which are circled.

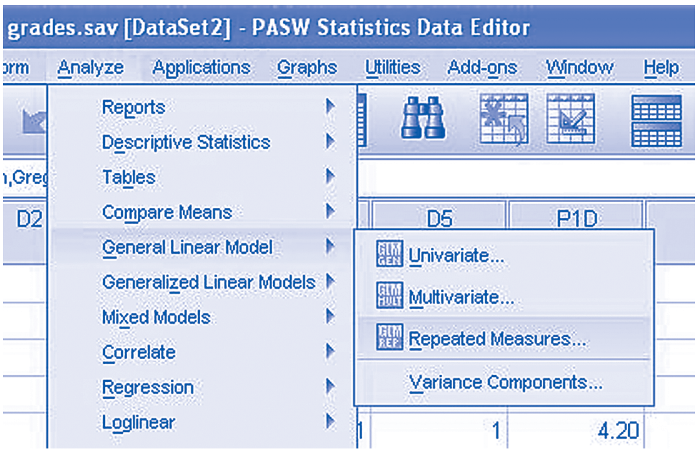
The first circled mean difference, which looks at the mean difference between whites and blacks, is 6170.21. This means that the average income for whites is $6170.21 greater than the average income for blacks. This result was significant at the $p < .001$ level. Next, the difference in income between whites and those of other race was found to be significant at the $p < .001$ level. Here, the mean income for whites was, on average, $2522.61 greater than that of members of other races. Finally, the difference in income between members of other races and blacks was found to be significant at the $p < .001$ level. In this case, the mean income for members of other races was, on average, $3647.60 greater than the average income for blacks.

Our results can be stated in the following way: A factorial ANOVA found a significant difference in income based on both race and gender, $F(5, 27157) = 573.72$, $p < .001$. Specifically, males were found to have significantly higher incomes than females, $F(1, 27157) = 383.95$, $p < .001$. Also, income was found to vary significantly based on race, $F(2, 27157) = 117.93$, $p < .001$. A Games-Howell post hoc test found that income for whites was significantly higher than that of blacks and those of other race, while the mean income for members of other races was significantly greater than the average income for blacks. Finally, a significant interaction between race and gender was found, $F(2, 27157) = 55.09$, $p < .001$.

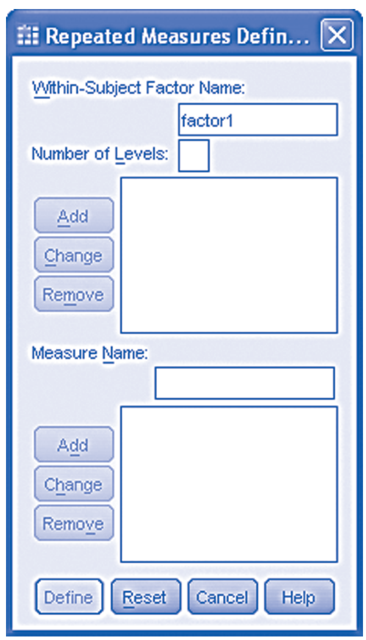
Repeated Measures ANOVA: IBM SPSS

Repeated measures ANOVAs are used when your dependent variable consists of a measure that was recorded or measured at several points in time. For example, if you had a set of two or more exam grades for a set of respondents, these data, along with one or more independent predictor variables, could be analyzed using a repeated measures ANOVA. This is the example that I'll be using in this section. A repeated measures ANOVA could also be used in other situations, for example, if you had a measure for respondents that was taken before and after some medical treatment. Using a repeated measures ANOVA, you can also include predictor variables such as sex and age.

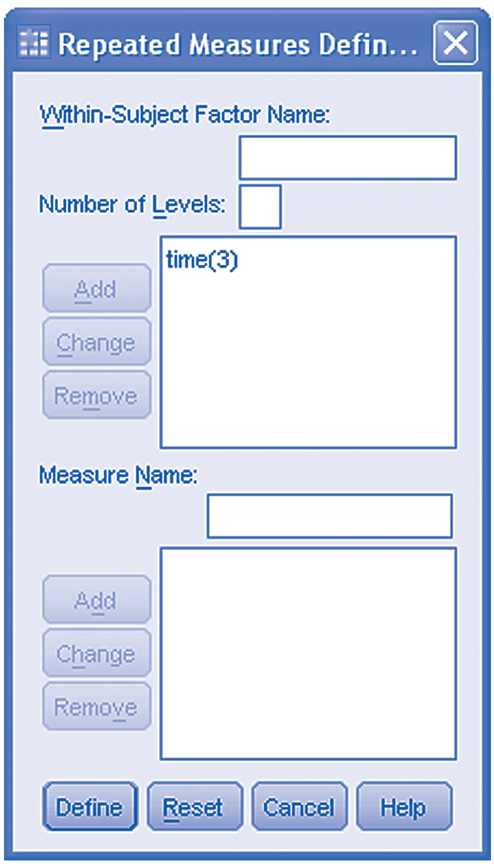
To run a repeated measures ANOVA, first make the following menu selection:



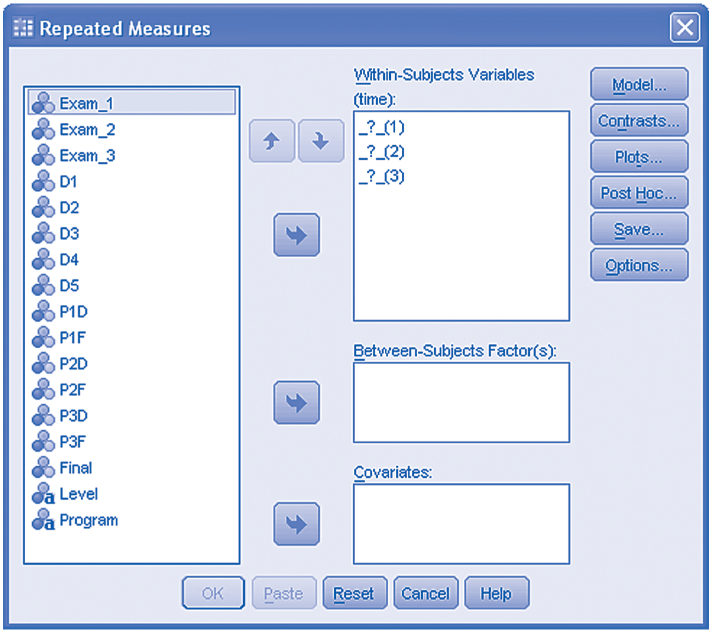
This will open the following dialog box:



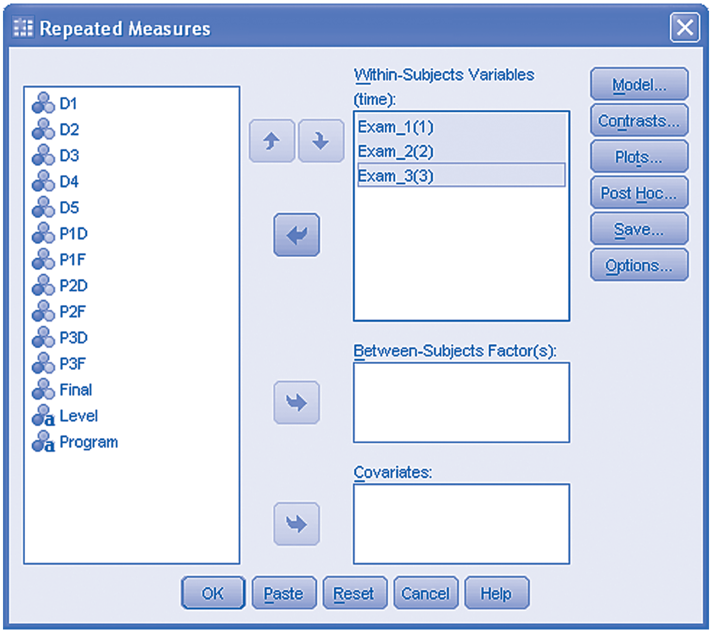
The data that I am using in this example consist of three exam scores in a sample of 37 students. The dependent variable consists of the three exam scores, while I will include year in college, categorized as Freshman, Sophomore, Junior, and Senior, as the independent variable. In the dialog box just presented, I will rename the *Within-Subject Factor Name* as simply *time*. Next, I will specify it as having three levels, as we have three separate exam scores. Finally, I will click *Add* under *Number of Levels*. When finished, the dialog box will look as follows:



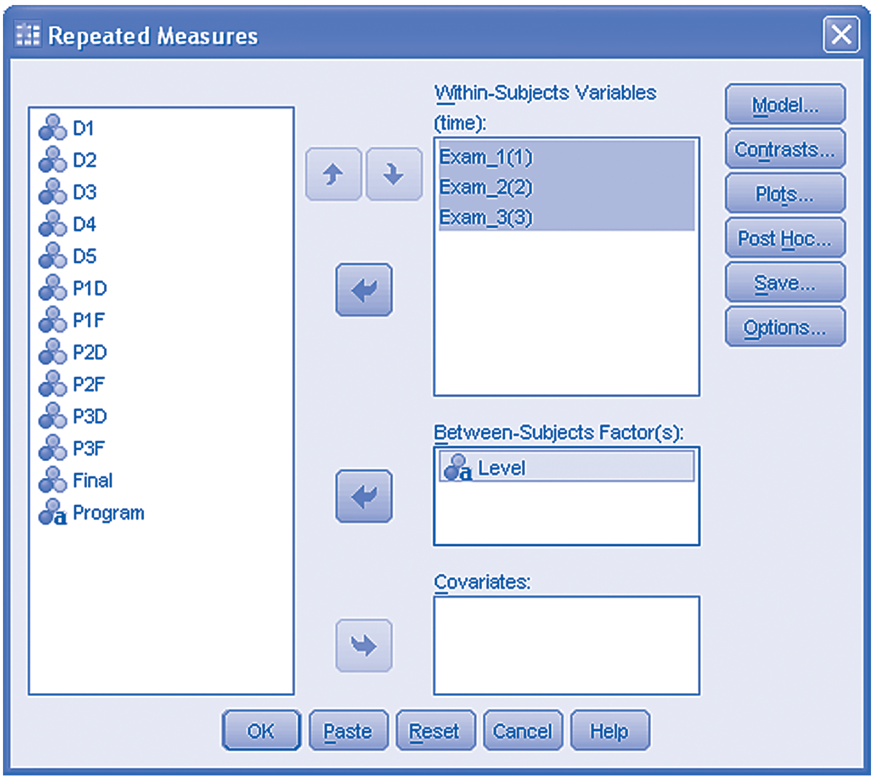
Next, we can click *Define*. This will reveal the following dialog box:



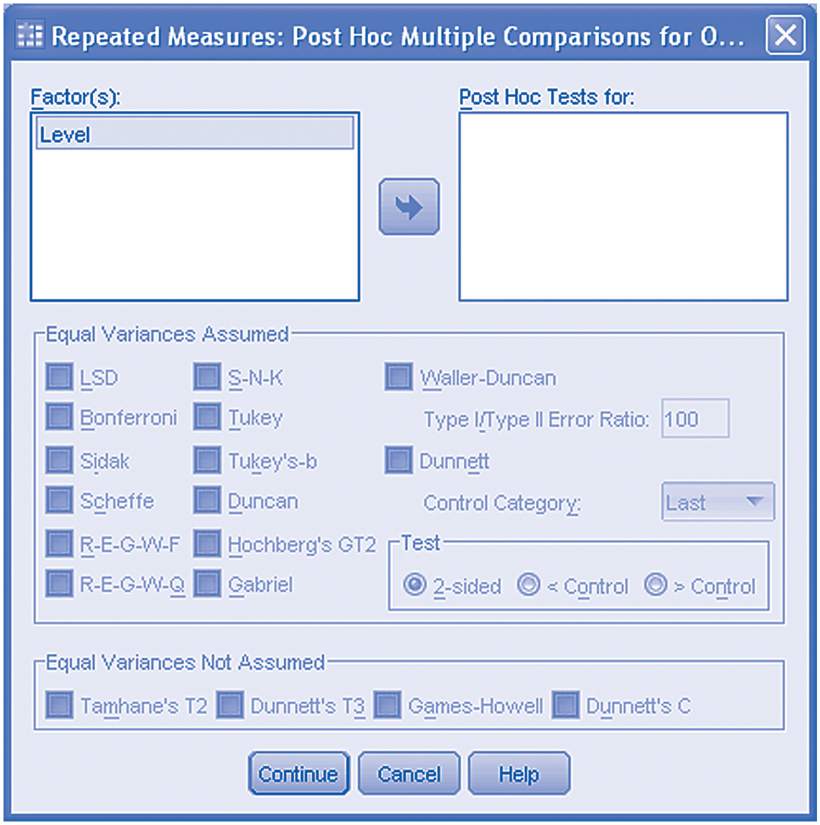
Here, we will begin by selecting the three exam scores, named *Exam_1*, *Exam_2*, and *Exam_3* and move them to the *Within-Subjects Variables (time)* box. After this step, the dialog box will look as follows:



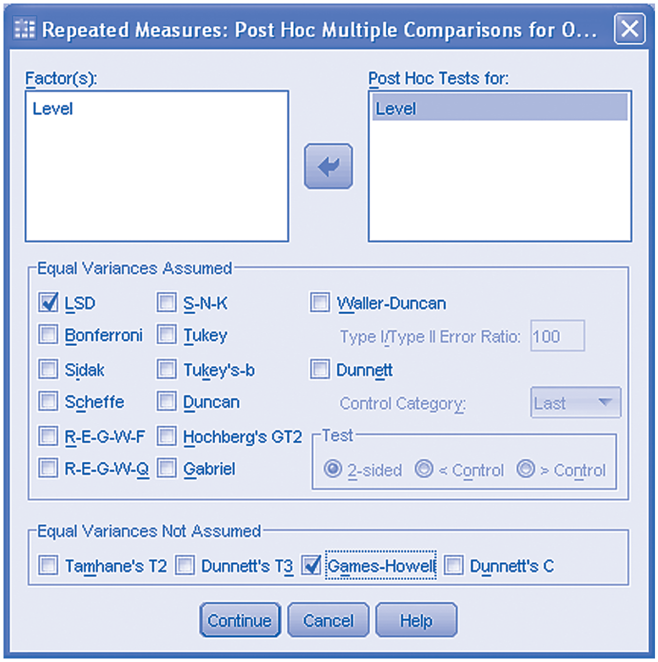
Next, we will specify our independent predictor variable, year at college. Any categorical predictor variables included in the repeated measures ANOVA will be included in the *Between-Subjects Factor(s)* box. In this example, we only have year at college as a categorical predictor variable, which is named *level* in this data set. After selecting this variable and moving it into the appropriate box, our dialog box will look as follows:



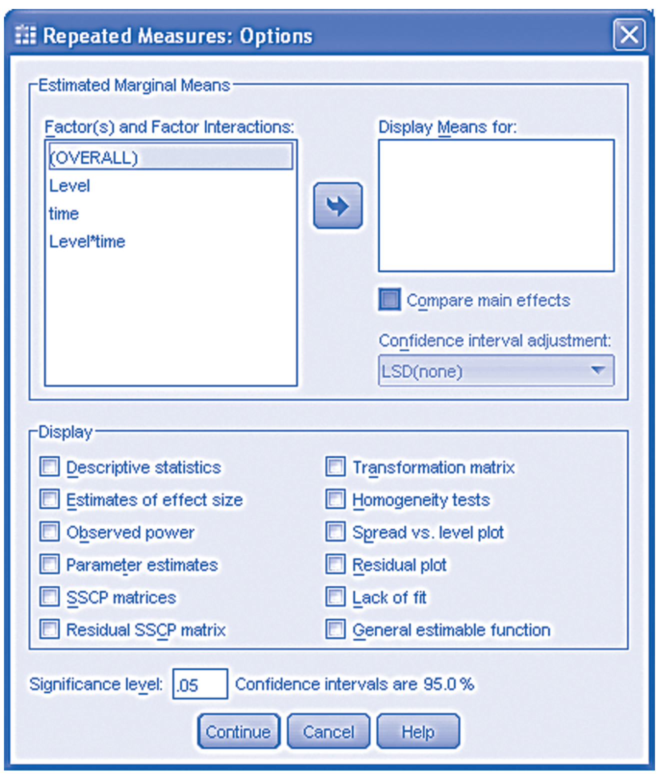
Next, let's click on the *Post Hoc* option, so we can specify post hoc tests for this ANOVA. This will allow us to see whether there are differences in exam scores between each category of year at college. For example, it will tell us whether Seniors have higher exam scores compared with Freshman. A post hoc test is needed here as the ANOVA will only tell you whether there are significant differences overall. The initial dialog box that you'll see when you first select this option is presented here:



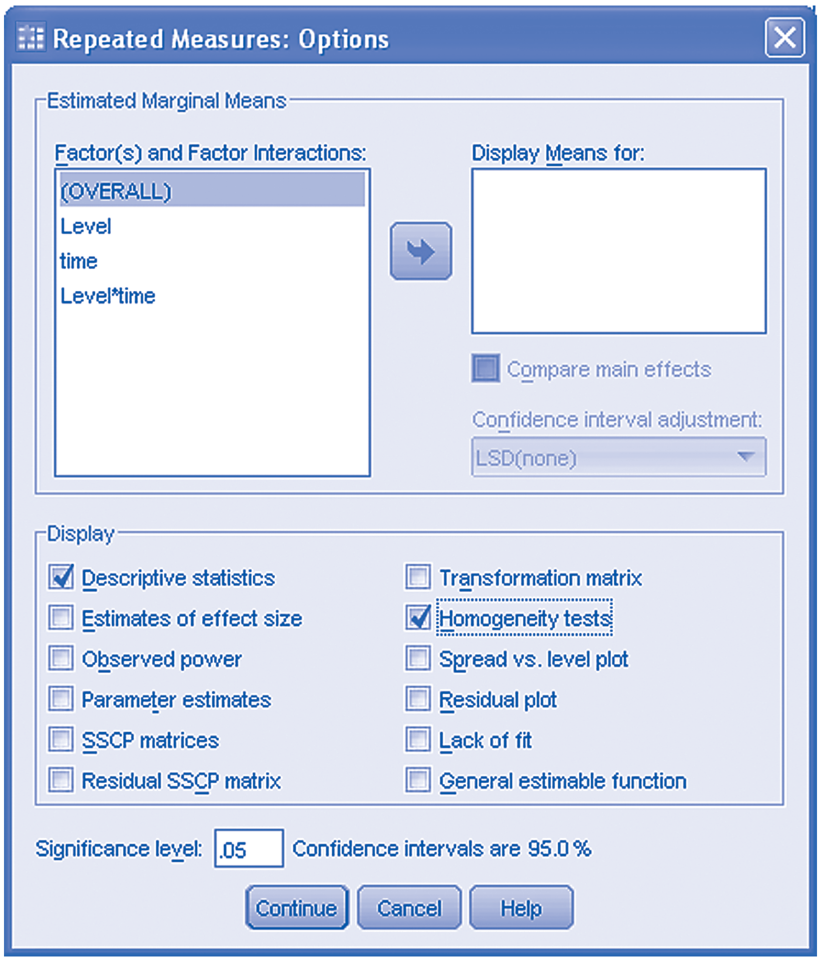
Next, we will simply select our *level* variable and move it to the right. Then, we will select the post hoc tests desired. Here, I will select both the LSD post hoc test as well as the Games-Howell post hoc test. The LSD post hoc test is less conservative while the Games-Howell post hoc test is more conservative. With more conservative tests, you are less likely to find a significant result, while their stricter standards mean that you're less likely to find a "false positive," or a result that is reported to be significant by SPSS, which in actuality is not. The LSD test incorporates the assumption that the variance in your dependent variable is approximately equal across the different categories of your independent variable, while the Games-Howell test does not. In this example, it would be assumed that variances in test scores are relatively the same regardless of the respondent's year of college. We will be testing this assumption which will determine which of these two post hoc tests we end up using in our analysis. After making our selections, our dialog box will appear as follows:



Next, click *Continue*. Then click *Options*. This opens the following dialog box:



Here, I will select *Descriptive statistics* and *Homogeneity tests*. The *Descriptive statistics* option will give us the mean of exam scores by year at college, while the *Homogeneity tests* option will test the assumption of equal variances. After making these selections, your dialog box should look as follows:



Finally, we can click *Continue* and *OK* to run the analysis.

Instead of presenting the results all at once, I will go through the tables a few at a time as running a repeated measures ANOVA in SPSS results in a large set of tables. The first three tables of the output are presented here:

Within-Subjects Factors

Measure:MEASURE_1

time	Dependent Variable
1	Exam_1
2	Exam_2
3	Exam_3

Between-Subjects Factors

		N
Level	Freshman	2
	Junior	8
	Senior	7
	Sophomore	20

Descriptive Statistics

	Level	Mean	Std. Deviation	N
Exam_1	Freshman	70.00	14.142	2
	Junior	91.75	5.258	8
	Senior	83.71	12.842	7
	Sophomore	83.10	12.226	20
	Total	84.38	11.910	37
Exam_2	Freshman	81.50	7.778	2
	Junior	93.00	3.780	8
	Senior	87.29	10.781	7
	Sophomore	84.10	13.242	20
	Total	86.49	11.423	37
Exam_3	Freshman	70.00	2.828	2
	Junior	84.50	6.568	8
	Senior	82.29	4.536	7
	Sophomore	70.40	25.111	20
	Total	75.68	19.695	37

The first table, "Within-Subjects Factors," simply presents the different measures of the dependent variable that we had specified. Here, you can see that we have simply specified the three different exam scores as the dependent variable in this repeated measures ANOVA. Next, the "Between-Subjects Factors" table presents the number of respondents for each category of our independent variable. The third table, titled "Descriptive Statistics," presents mean scores for each exam separately by year at college. For example, the first row presents the mean score on Exam 1 for freshmen, which is 70.00. The final row presents the mean score on Exam 3 for all respondents, which is 75.68.

Here are the next two tables of the output:

Box's Test of Equality of Covariance Matrices[a]

Box's M	46.536
F	3.151
df1	12.000
df2	1470.214
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Level Within Subjects Design: time

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	.108	1.941[a]	2.000	32.000	.160
	Wilks' Lambda	.892	1.941[a]	2.000	32.000	.160
	Hotelling's Trace	.121	1.941[a]	2.000	32.000	.160
	Roy's Largest Root	.121	1.941[a]	2.000	32.000	.160
time * Level	Pillai's Trace	.107	.625	6.000	66.000	.710
	Wilks' Lambda	.894	.615[a]	6.000	64.000	.717
	Hotelling's Trace	.117	.605	6.000	62.000	.725
	Roy's Largest Root	.102	1.123[b]	3.000	33.000	.354

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + Level
Within Subjects Design: time

The first table presented here, titled "Box's Test of Equality of Covariance Matrices" includes a calculation of Box's *M* statistic and its significance. The statistic is used to test the assumptions of the multivariate model, which will be explained shortly. If the significance of the statistic is less than .05, it means that the assumptions of the multivariate model have been violated, and therefore, the multivariate model should not be used. Here, you can see that the probability level is less than .001, which means that the assumptions of the multivariate model have been violated.

The next table, titled "Multivariate Tests," presents the results of the repeated measures ANOVA. In short, "multivariate" means that you are incorporating more than one predictor variable, while "univariate" means that you are incorporating only one predictor. In this example, both *level* (year at college) and *time* are included, making this a multivariate model. You can see that for each variable or interaction effect included, SPSS gives you four different versions of the *F* test. Wilks's Lambda is very commonly used, so I'll focus on that version here. As mentioned in the previous paragraph, as the assumptions of the multivariate model have been violated, you would prefer not to focus on the multivariate model. However, I will explain the results for your understanding. We can see that the Wilks's Lambda *F* test calculated an *F* statistic for *time* of 1.941 with a *p* value of .160. This means that in this multivariate model, test scores were not found to significantly vary based on time. In regard to the time * level interaction effect, the calculated *F* statistic using the Wilks's Lambda *F* test was .615 with a *p* level of .717. This means that in this multivariate model, the effect of year at college on test scores did not vary significantly based on time. Alternatively, you could state that the effect of time on test scores did not vary significantly based on year at college.

The next two tables are presented here:

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subject...	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.600	16.323	2	.000	.715	.806	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + Level Within Subjects Design: time

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	918.973	2	459.486	2.658	.078
	Greenhouse-Geisser	918.973	1.429	643.082	2.658	.097
	Huynh-Feldt	918.973	1.611	570.295	2.658	.090
	Lower-bound	918.973	1.000	918.973	2.658	.113
time * Level	Sphericity Assumed	509.048	6	84.841	.491	.813
	Greenhouse-Geisser	509.048	4.287	118.741	.491	.755
	Huynh-Feldt	509.048	4.834	105.301	.491	.776
	Lower-bound	509.048	3.000	169.683	.491	.691
Error(time)	Sphericity Assumed	11407.943	66	172.848		
	Greenhouse-Geisser	11407.943	47.157	241.912		
	Huynh-Feldt	11407.943	53.176	214.531		
	Lower-bound	11407.943	33.000	345.695		

The first table, titled "Mauchly's Test of Sphericity" presents a test of the assumptions of the univariate model, the results of which are presented in the second table. The probability level, which is circled, is below .05, which means that the assumption of sphericity has been violated. However, this does not prevent us from using the results of the univariate model. The final three columns of the table present three corrections to the calculated *F* statistic. The Huynh-Feldt correction is somewhat less conservative than the others and is what I will focus on here.

The second table, titled "Tests of Within-Subjects Effects," presents the effect of time on test scores. As discussed in the previous paragraph, the Huynh-Feldt correction will be used here. The effect of time on test scores was not found to be significant, having an *F* value of 2.658 with a probability level of .090. The interaction between time and year at college (*level*) was also not found to be significant, having an *F* value of .491 with a probability level of .776. I will skip the next table of the output, titled "Tests of Within-Subjects Contrasts," as it is not commonly used.

The next two tables are presented here:

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Exam_1	2.076	3	33	.122
Exam_2	1.532	3	33	.225
Exam_3	1.628	3	33	.202

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Level
Within Subjects Design: time

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	392737.160	1	392737.160	1376.014	.000
Level	2432.320	3	810.773	2.841	.053
Error	9418.743	33	285.416		

The first table, titled "Levene's Test of Equality of Error Variances," tests the assumption that the variances in test scores are equal across the categories of the independent variable, which is year at college in this example. This is important for the post hoc test that will be presented shortly, as a number of post hoc tests assume that these variances are equal. As you can see, the probability levels, which are circled, are not significant at the .05 level, which means that this assumption has not been violated and that we can use post hoc tests that assume the equality of variances.

The second table, titled "Tests of Between-Subjects Effects," tests the effect of our independent variable, *level* or year at college, on exam scores. As you can see, this effect approaches significance with an *F* value of 2.841 and a probability level of .053.

The final table, presenting the results of our post hoc tests, is presented here:

Post Hoc Tests

Level

Multiple Comparisons

Measure:MEASURE_1

	(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LSD	Freshman	Junior	-15.92*	7.711	.047	-31.61	-.23
		Senior	-10.60	7.821	.185	-26.51	5.32
		Sophomore	-5.37	7.234	.463	-20.08	9.35
	Junior	Freshman	15.92*	7.711	.047	.23	31.61
		Senior	5.32	5.048	.299	-4.95	15.59
		Sophomore	10.55*	4.080	.014	2.25	18.85
	Senior	Freshman	10.60	7.821	.185	-5.32	26.51
		Junior	-5.32	5.048	.299	-15.59	4.95
		Sophomore	5.23	4.283	.231	-3.49	13.94
	Sophomore	Freshman	5.37	7.234	.463	-9.35	20.08
		Junior	-10.55*	4.080	.014	-18.85	-2.25
		Senior	-5.23	4.283	.231	-13.94	3.49
Games-Howell	Freshman	Junior	-15.92	2.316	.113	-46.04	14.21
		Senior	-10.60	2.870	.107	-25.11	3.92
		Sophomore	-5.37	3.527	.480	-17.49	6.75
	Junior	Freshman	15.92	2.316	.113	-14.21	46.04
		Senior	5.32	2.053	.116	-1.21	11.85
		Sophomore	10.55*	2.901	.007	2.49	18.61
	Senior	Freshman	10.60	2.870	.107	-3.92	25.11
		Junior	-5.32	2.053	.116	-11.85	1.21
		Sophomore	5.23	3.360	.421	-4.03	14.49
	Sophomore	Freshman	5.37	3.527	.480	-6.75	17.49
		Junior	-10.55*	2.901	.007	-18.61	-2.49
		Senior	-5.23	3.360	.421	-14.49	4.03

Based on observed means.
The error term is Mean Square(Error) = 95.139.

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

As mentioned earlier, Levene's test of the equality of variances found that the assumption that the variances are equal was not violated. This means that we can use a post hoc test that assumes equal variances. The LSD post hoc test, presented in this table, assumes equal variances and will be used in this analysis. As mentioned earlier in this chapter, as every possible group comparison is included in the table of the post hoc test results, all comparisons will appear twice. For example, the "Junior versus Freshman" comparison, with a mean difference of 15.92 (circled), is also presented in the first row of the table as the "Freshman versus Junior" comparison, with a mean difference of –15.92. A good way to make sure that you do not report the same comparison twice is to simply focus only on significant comparisons that have a positive mean difference.

The first significant comparison was that of Junior and Freshman. Juniors were found to have test scores that were, on average, 15.92 points higher than that of Freshman. This effect was found to be significant, having a probability level of below .05. The second significant comparison, also circled, was between Junior and Sophomore. Juniors were found to have test scores that were, on average, 10.55 points higher than that of Sophomore. This effect was found to be significant, also having a probability level of below .05.

Our results can be stated in the following way:

A repeated measures ANOVA was conducted on a series of three exam grades with year at college (Freshman, Sophomore, Junior, and Senior) as the independent predictor. The multivariate model will not be used as Box's M test found significant variability in the observed **covariance** matrices of the dependent variables across groups. Mauchly's test of sphericity was found to be significant at the .05 alpha level; hence, the Huynh-Feldt adjustment will be used in the univariate model. Neither the effects of time nor the interaction between time and year at college were found to significantly predict exam grades. However, the effect of year at college approached significance, $F(3, 33) = 2.84$, $p = .053$.

An LSD post hoc test was used to analyze the differences in exam grades based on year at college. Significant differences in exam scores were found between Junior and Freshman and between Junior and Sophomore. Juniors were found to have test scores that were, on average, 15.92 points higher than that of Freshman, $p < .05$. Juniors were also found to have test scores that were, on average, 10.55 points higher than that of Sophomores, $p < .05$.

ANOVA Syntax Summary

One-Way ANOVA

```
ONEWAY educ BY race
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS
/POSTHOC=LSD GH ALPHA(.05).
```

The general format being as follows:

```
ONEWAY [Dependent variable] BY [Independent
variable]
/STATISTICS [Options]
/MISSING ANALYSIS
```

```
/POSTHOC=[Post hoc tests] ALPHA([Alpha or
  probability level]).
```

Factorial ANOVA

```
UNIANOVA realrinc BY race sex
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/POSTHOC=race(LSD GH)
/PRINT=HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=race sex race*sex.
```

The general format being as follows:

```
UNIANOVA [Dependent variable] BY [List of
  independent variables]
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/POSTHOC=[Independent variables to include in
the post hoc test] ([Post hoc tests])
/PRINT=[Options]
/CRITERIA=ALPHA([Alpha or probability level])
/DESIGN=[Design of model].
```

Repeated Measures ANOVA

```
GLM Exam_1 Exam_2 Exam_3 BY Level
/WSFACTOR=time 3 Polynomial
/METHOD=SSTYPE(3)
/POSTHOC=Level(LSD GH)
/PRINT=DESCRIPTIVE HOMOGENEITY
/CRITERIA=ALPHA(.05)
/WSDESIGN=time
/DESIGN=Level.
```

The general format being as follows:

```
GLM [Dependent measures] BY [Independent
  variable]
/WSFACTOR=[Repeated measures "factor" name (can
  specify any name)] [Number of times the
  measure is repeated] Polynomial
```

```
/METHOD=SSTYPE(3)
/POSTHOC=[Independent variables to include in
  the post hoc test] ([ Post hoc tests])
/PRINT=[Options]
/CRITERIA=ALPHA([Alpha or probability level])
/WSDESIGN=[Repeated measures "factor" name]
/DESIGN=[Design of model] .
```

SECTION 3: STATA

Pearson's *r*: Stata

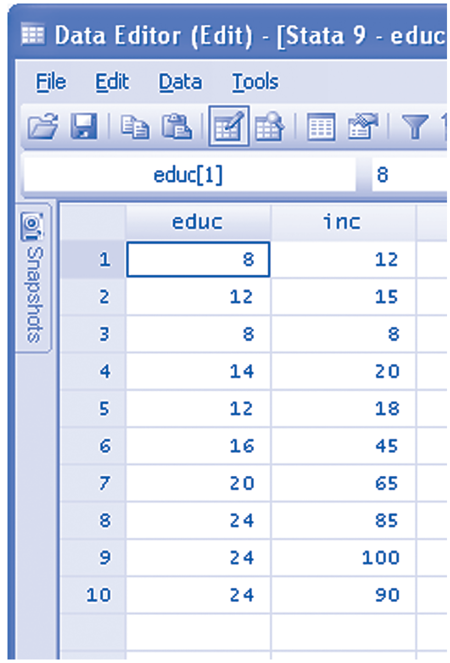
Calculating Pearson's *r* in Stata is quite simple. In this section, I will use the same data as were presented in the first section on Pearson's *r*. First, I will create two new variables, one for income and one for education, by entering the following commands:

```
gen inc=.
gen educ=.
```

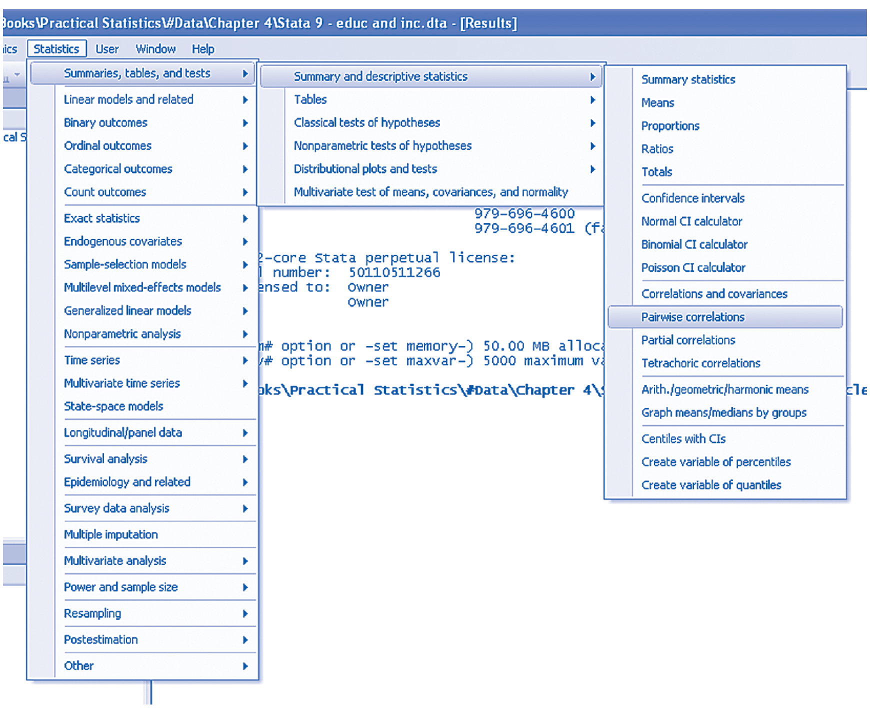
Next, I will enter the data displayed in the table below using the data editor by entering the command `ed`, which stands for edit.

Years of Education (x)	Income (in Thousands of $) (y)
8	12
12	15
8	8
14	20
12	18
16	45
20	65
24	85
24	100
24	90

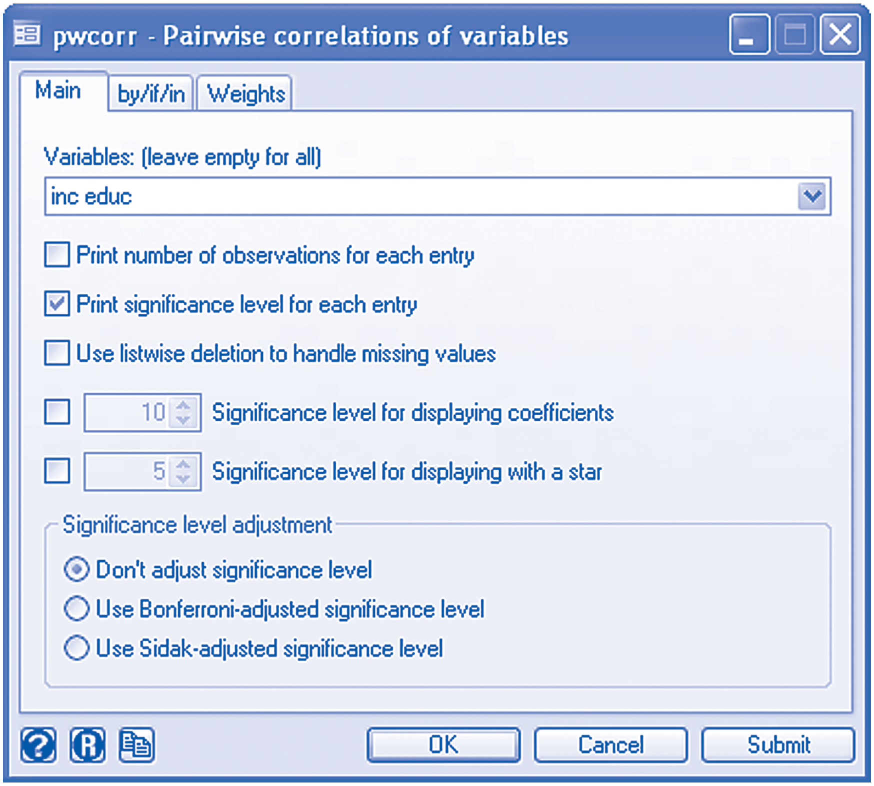
When you have finished entering data, your data editor should look like this:



After closing the editor, we can make the following menu selection:



This will open the following dialog box:




```
. pwcorr inc educ, sig

             |      inc     educ
-------------+------------------
         inc |   1.0000
             |
        educ |   0.9743   1.0000
             |   0.0000
```

As you can see, I have listed the income and education variables under the *Variables* entry. Also, I have selected the *Print significance level for each entry* option so that the *p* level is included in our results. Clicking *OK* will give us this result:

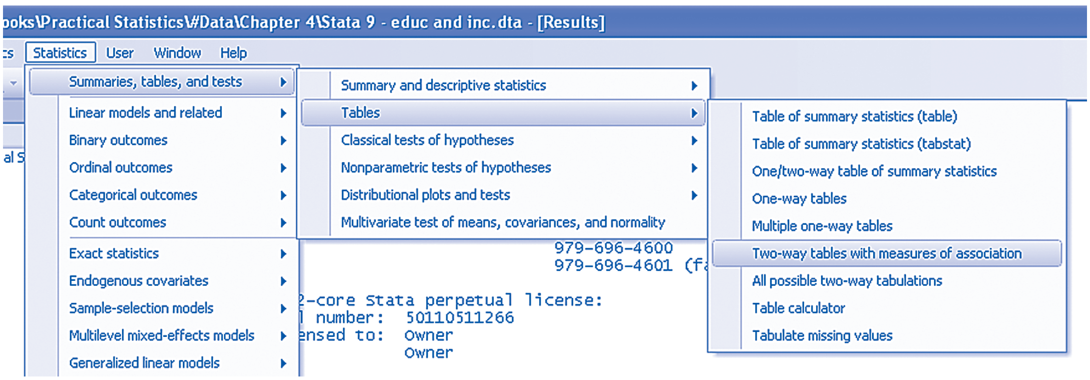
Comparing this result from the one we obtained in the initial section on Pearson's *r*, we can see that the calculated correlation coefficient is identical. Secondly, we have generated the syntax that is used to determine the correlation coefficient between two variables in Stata:

```
pwcorr inc educ, sig
```

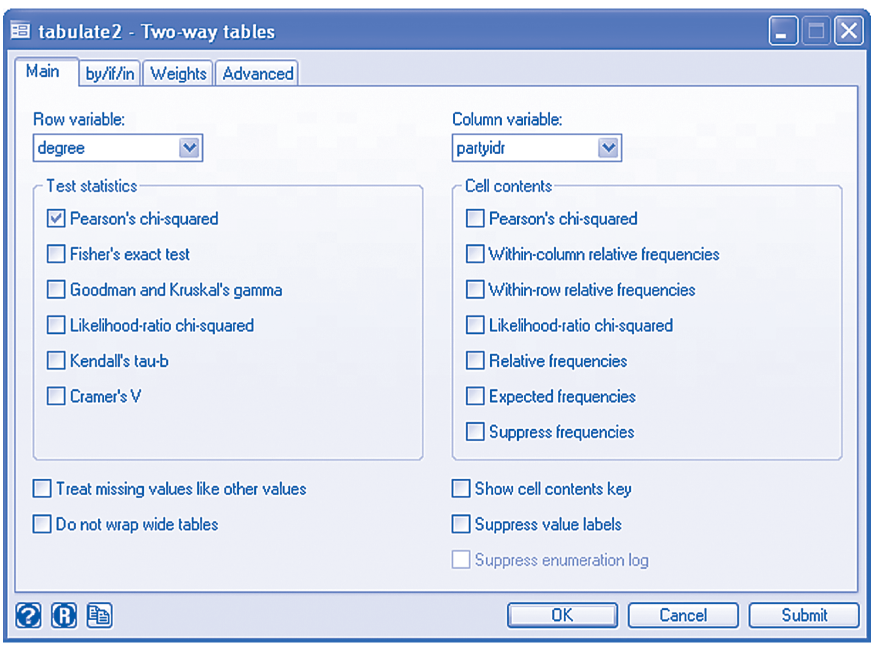
Specifying the sig option tells Stata to include the significance level of the correlation coefficient in the results. Here, our *p* level is listed as "0.0000," which simply means that our true *p* level is less than .0001. Stata incorrectly rounded our *p* level down to zero, while it can never be zero in actuality. Here, we could say the following: There is a statistically significant positive correlation between years of education and income ($r = .97$, $p < .001$).

Chi-Square: Stata

To calculate a chi-square statistic in Stata, first navigate to the following menu selection:



This will open the following dialog box:



Using the same example from the IBM SPSS section, I have selected *degree* under *Row variable* and the recoded *partyid* variable, which I have renamed *partyidr*, under *Column variable*. Clicking *OK* will give you the following output:

```
. tabulate degree partyidr, chi2

                      |         recode party id
               degree |  democrat  independe  republica |     Total
----------------------+---------------------------------+----------
less than high school |     6,334      1,698      2,832 |    10,864
          high school |    11,563      3,529      8,513 |    23,605
associate/junior coll |     1,045        303        825 |     2,173
           bachelor's |     2,618        601      2,812 |     6,031
             graduate |     1,431        286      1,072 |     2,789
----------------------+---------------------------------+----------
                Total |    22,991      6,417     16,054 |    45,462

          Pearson chi2(8) = 834.0072   Pr = 0.000
```

As you can see, both IBM SPSS and Stata calculated a Pearson chi-square value of 834.007 in this particular example. As you may notice, Stata puts the degrees of freedom on the last line right next to chi2, in parentheses. As you may remember, in this example, the degrees of freedom was 8. Stata also calculates the probability level as being less than .001. Here, we could say the following: There was a statistically significant relationship between highest degree completed and political party affiliation ($\chi^2 = 834.01$, $df = 8$, $p < .001$).

Finally, this is the corresponding Stata syntax for this particular example:

```
tabulate degree partyidr, chi2
```

t-Test: Stata

To run an independent samples *t*-test within Stata, we must first test whether our dependent variable has equal variances across groups. We will do this using Levene's test for the equality of variances as was reported in the previous section on running *t*-tests within IBM SPSS. In this example, we are testing differences in respondents' income based on sex. To test whether the variances in scores are equal across sex, we will use the following syntax command:

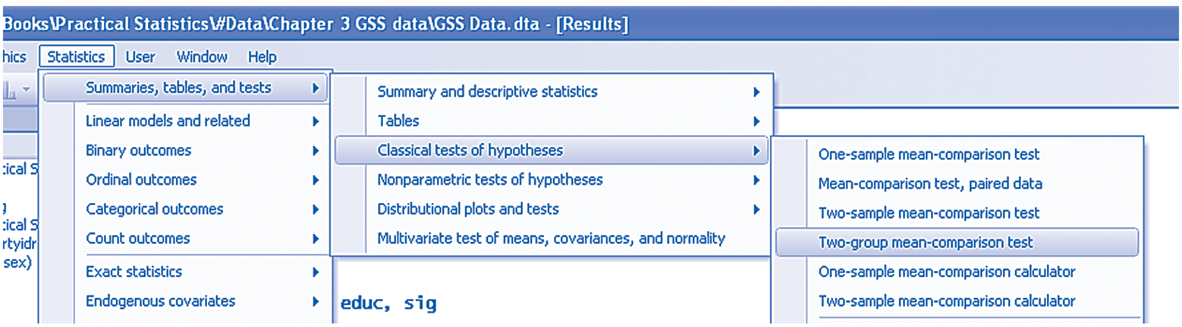
```
robvar realrinc, by(sex)
```

Typing this command into the command window and hitting enter would give you the following results:

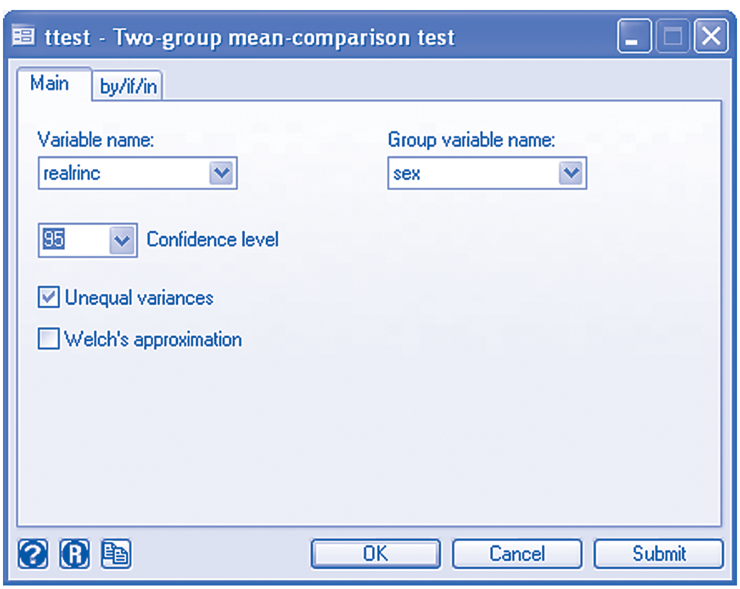
Look at the first line, under "w0." As you can see, the calculated *F* statistic for Levene's test for the equality of variances here in Stata is identical to the score calculated previously in IBM SPSS. Also, looking at the $Pr > F =$ entry for *w0*, you notice that this test is significant at the .05 level.

```
. robvar realrinc, by(sex)

            |          Summary of realrinc
        sex |        Mean   Std. Dev.       Freq.
------------+------------------------------------
          1 |    26983.16   24339.702      13862
          2 |   14811.216   14022.518      13301
------------+------------------------------------
      Total |   21022.882   20871.594      27163

W0  =  1318.0868    df(1, 27161)     Pr > F = 0.00000000

W50 =  1012.3438    df(1, 27161)     Pr > F = 0.00000000

W10 =  1054.4143    df(1, 27161)     Pr > F = 0.00000000
```

Next, we will run the actual *t*-test. First, choose the following menu selection:



This will load the following dialog box:



As you can see, I have specified *sex* under *Group variable name* (which will always contain the variable that you have two groups of) and *realrinc* under *Variable name* (the dependent, continuous variable). Because we know that the variances between groups are significantly different, I have also specified *Unequal variances* in this dialog box. Clicking *OK* will give us the following output:

```
. ttest realrinc, by(sex) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |   13862    26983.16    206.7294     24339.7    26577.94    27388.38
       2 |   13301    14811.22    121.5861    14022.52    14572.89    15049.54
---------+--------------------------------------------------------------------
combined |   27163    21022.88    126.6388    20871.59    20774.66     21271.1
---------+--------------------------------------------------------------------
    diff |            12171.94    239.8338                11701.85    12642.03
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =  50.7516
Ho: diff = 0                     Satterthwaite's degrees of freedom =  22324.9

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

We can see that the calculated *t* score is 50.7516 with 22324.9 degrees of freedom. For the probability level, we can look at the second entry at the bottom. Here, we can see that our results are significant at the $p < .0001$ probability level under the "Pr(|T| > |t|)" entry, which represents the *p* level for the two-tailed *t*-test. Stata's output also gives us the means of the two different groups, along with the number of observations and several other statistics. These results could be stated as follows: Males were found to have a significantly higher income as compared with female respondents ($t = 50.75$, $df = 22324.9$, $p < .001$).

This is the corresponding syntax:

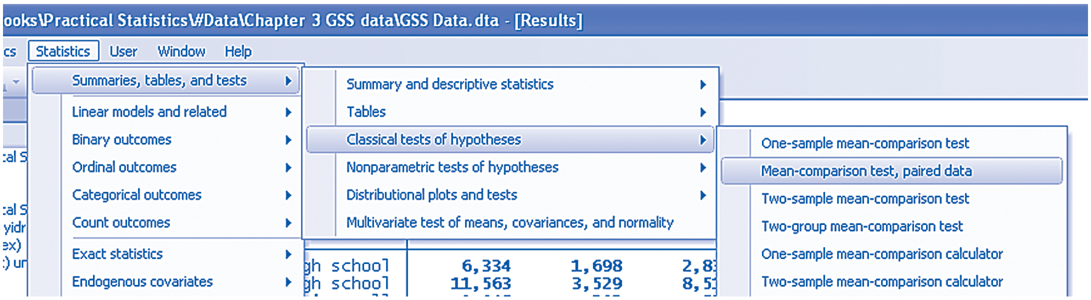
```
ttest realrinc, by(sex) unequal
```

And this syntax would be used if you assumed equal variances:

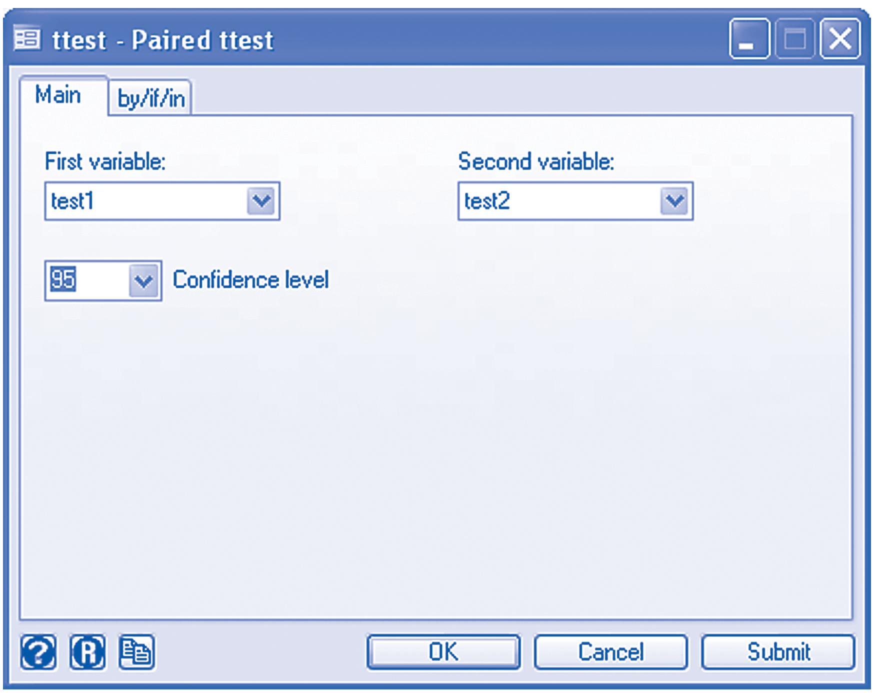
```
ttest realrinc, by(sex)
```

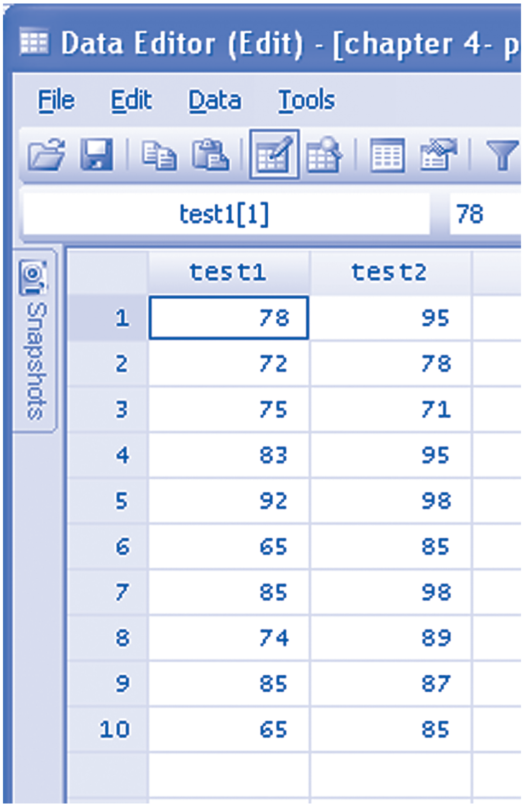
Now, let's use Stata to run a paired samples or dependent *t*-test. I will use the same example as used in the previous IBM SPSS example in which we had

10 respondents who took an exam at two different periods of time. I will also use the identical values that were used previously. First, navigate to the following menu selection:



This will reveal the following dialog box:





	test1	test2
1	78	95
2	72	78
3	75	71
4	83	95
5	92	98
6	65	85
7	85	98
8	74	89
9	85	87
10	65	85

Here, I have simply specified *test1* and *test2* as the two variables to be included in this paired *t*-test. These are the data I am using in this example:

Clicking *OK* will give us the following output:

```
. ttest test1 == test2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   test1 |      10        77.4    2.817406    8.909421    71.02658    83.77342
   test2 |      10        88.1    2.802578     8.86253    81.76013    94.43987
---------+--------------------------------------------------------------------
    diff |      10       -10.7    2.525646    7.986795   -16.41341   -4.986591
------------------------------------------------------------------------------
     mean(diff) = mean(test1 - test2)                             t =  -4.2365
 Ho: mean(diff) = 0                              degrees of freedom =        9

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0011       Pr(|T| > |t|) = 0.0022          Pr(T > t) = 0.9989
```

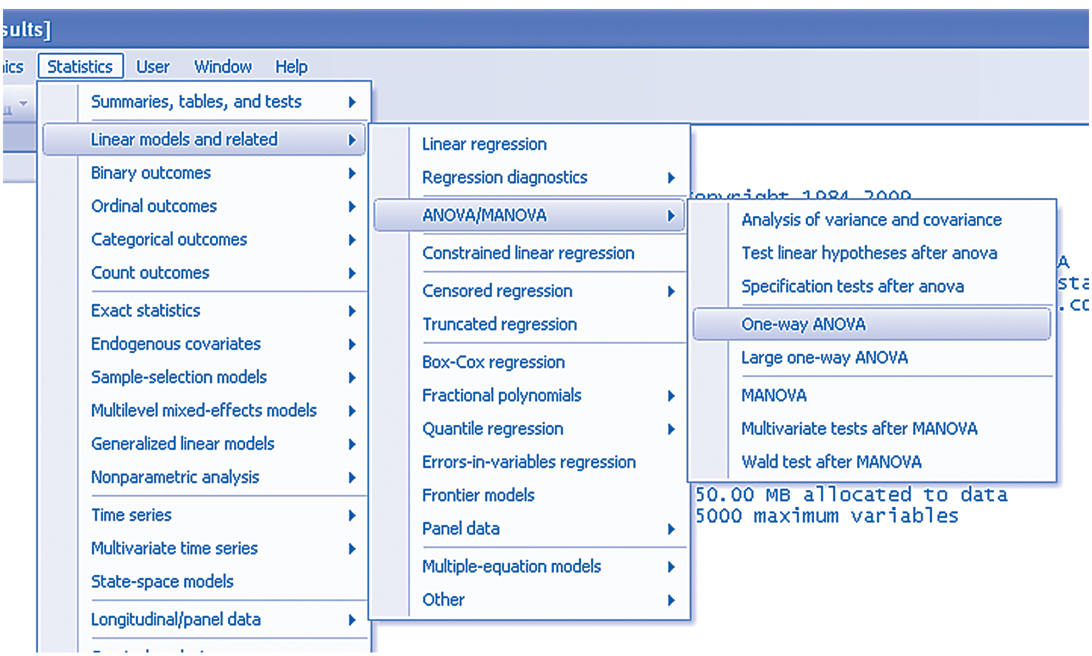
Here, we see that for this **paired samples *t*-test**, Stata has calculated a *t* value of –4.2365 with 9 degrees of freedom. By again looking at the middle entry at the bottom of the output, which is used for a two-tailed *t*-test, we see that this is significant at the .0022 probability level. You could report this result in the following way: "Scores on Test 2 were found to be significantly different from scores on Test 1 ($t = -4.24$, $df = 9$, $p < .01$). Specifically, the mean of Test 2 scores was 10.7 points higher than the mean of Test 1 scores." This second sentence is constructed using the values under the *Mean* column as well as the value under the *diff* row, which represents the difference between our two variables (*test1* and *test2*).

Finally, the equivalent syntax is simply this:

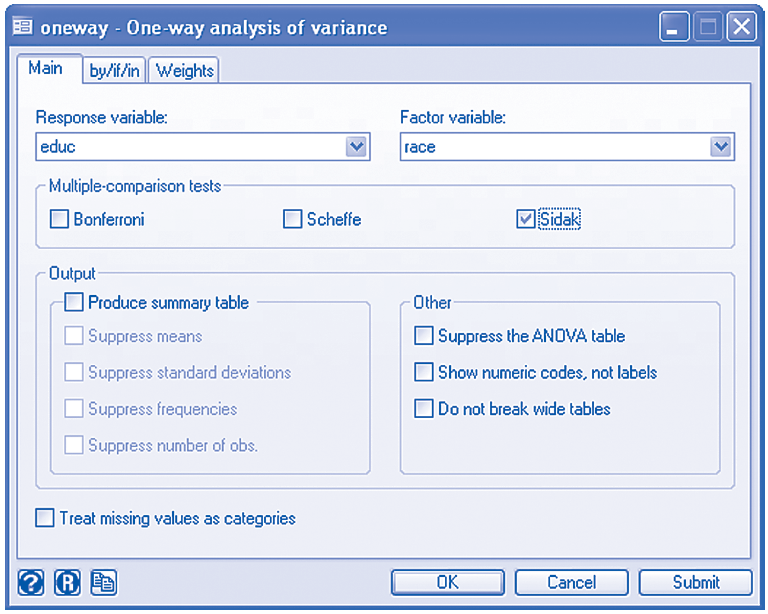
```
ttest test1 == test2
```

One-Way ANOVA: Stata

To run an ANOVA within Stata, first navigate to the following menu selection:



This will bring up the following dialog box:



Using the same example as used previously, I have specified *educ* as the *Response variable* and *race* as the *Factor variable*. I also specified that the Sidak post hoc analysis (here referred to as "Multiple-comparison tests") be run. As you can see, Stata is much more limited than SPSS in regard to the number of post hoc tests it supports. Clicking *OK* will give you the following results:

```
. oneway educ race, sidak

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      6898.37418      2   3449.18709    349.10     0.0000
 Within groups      458112.554  46366   9.8803553
------------------------------------------------------------------------
    Total           465010.928  46368   10.0287036

Bartlett's test for equal variances:  chi2(2) = 108.9791  Prob>chi2 = 0.000

                         Comparison of educ by race
                                  (Sidak)
Row Mean-|
Col Mean |          1           2
---------+----------------------
       2 |   -1.11246
         |      0.000
         |
       3 |     .14714      1.2596
         |      0.181       0.000
```

As you can see, Stata has calculated the *F* statistic to be 349.10 with the degrees of freedom of 2 and 46,366, significant at the $p < .001$ probability level. In the Sidak post hoc table shown below the ANOVA results, we see that only Groups 1 and 2 and Groups 2 and 3 are significantly different from each other at the .05 probability level, just as we found within IBM SPSS. We can tell this from the significance levels. Here, there are only two comparisons in which the probability level is below .05: the comparison between Groups 1 and 2 (whites and blacks, respectively) and between Groups 2 and 3 (blacks and those of other race, respectively). Our results can be stated in the following way:

> A significant difference in the level of education between whites, blacks, and those of other races was found ($F(2, 46366) = 349.10$, $p < .001$). Specifically, a Sidak post hoc test found the mean income for whites to be significantly different from that of blacks, with whites having greater incomes, on average, than blacks. Also, the mean income for blacks was found to significantly differ from those of other races, with members of other races having higher incomes as compared with blacks.

Our value of –1.11 at the top of the table represents the mean of blacks' education (coded 2) minus the mean of whites' education (coded 1), where education was coded in years. This value of –1.11 means that the mean of years of education for blacks is less than that of whites, as it is negative. The value of 1.2596 in the lower right-hand cell of the table represents the mean of education for those of other race (coded 3) minus the mean of blacks' education (coded 2). The values will always represent the mean for the row (in this case, 3) minus the mean for the column (in this case, 2).

We could check this using the following syntax:

```
. tabstat educ, by(race)

Summary for variables: educ
     by categories of: race

    race |      mean
---------+----------
       1 |  12.75507
       2 |  11.64261
       3 |  12.90221
---------+----------
   Total |  12.60765
--------------------
```

Here, we see that the mean years of education for blacks, coded 2, is less than that of whites, coded 1.

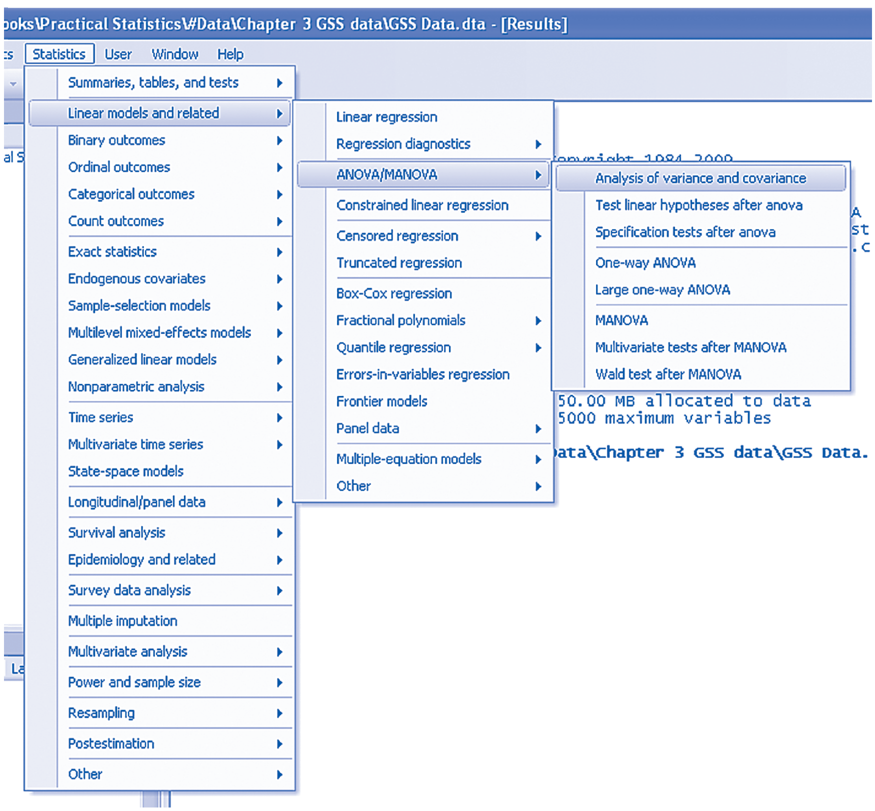
This is the corresponding syntax for the ANOVA:

```
oneway educ race, sidak
```

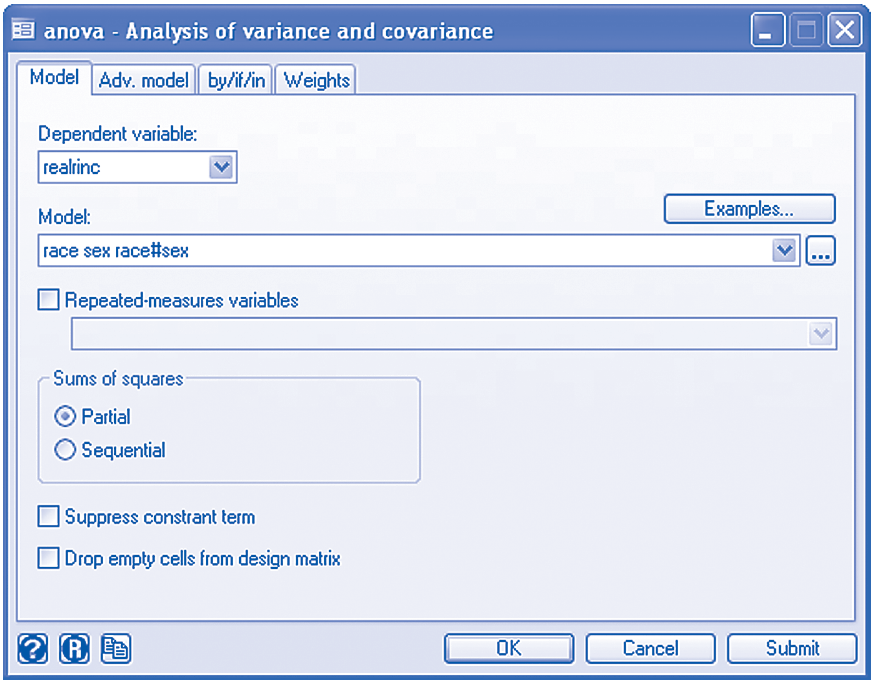
Factorial ANOVA: Stata

While one-way ANOVAs only include one categorical independent or predictor variable, factorial ANOVAs include more than one. In the example presented in the previous section, a one-way ANOVA was used as there was only one independent or predictor variable, race of the respondent. In this example, I will incorporate both race of the respondent as well as the respondent's gender in a factorial ANOVA that includes the respondent's income as the dependent variable. This example will also use data from the GSS.

To begin, first make the following menu selection:



This will open the following dialog box:



As you can see, I have specified the variable *realrinc*, a measure of the respondent's annual income, as the dependent variable. I have also specified the model to include two independent variables, the respondents' race and sex. Also, I have included as a term *race#sex*, which is the interaction between these two variables. It is possible that the effect of race on income varies by gender or likewise that the effect of gender on income varies by race. The inclusion of this interaction effect will test whether this is the case. After clicking *OK*, you will see the following results:

```
. anova realrinc race sex race#sex

                           Number of obs =    27163     R-squared     =  0.0955
                           Root MSE      =  19851.4     Adj R-squared =  0.0954

                  Source | Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model | 1.1304e+12     5  2.2609e+11     573.72     0.0000
                         |
                    race | 9.2944e+10     2  4.6472e+10     117.93     0.0000
                     sex | 1.5131e+11     1  1.5131e+11     383.95     0.0000
                race#sex | 4.3417e+10     2  2.1708e+10      55.09     0.0000
                         |
                Residual | 1.0702e+13 27157   394077231
              -----------+----------------------------------------------------
                   Total | 1.1832e+13 27162   435623423
```

As you can see under the “Model” row, this ANOVA had a calculated *F* statistic of 573.72, with a probability level of less than .001. The effect of race, sex, and the interaction between race and sex were all found to be significant. First, the effect of race on income was found to be significant, having an *F* statistic of 117.93 with a *p* level of less than .001. Next, the effect of sex on income was also found to be significant, having an *F* statistic of 383.95 with a *p* level of less than .001. Finally, the interaction between race and sex was found to be significant, having an *F* statistic of 55.09 with a *p* level of less than .001. This means that the effect of race on income significantly varies by sex. Alternatively, you could state that the effect of sex on income varies significantly by race. For example, this would be the case if race was an important predictor of income for males but not for females. Likewise, this would be the case if males have higher incomes than females for whites but if females had higher incomes than males for blacks. In essence, the significant interaction effect in this example means that the effect of one of the independent variables on the dependent variable varies significantly depending on the level of the second independent variable. Interaction effects can clearly be trickier to deal with and can take some additional time to fully understand.

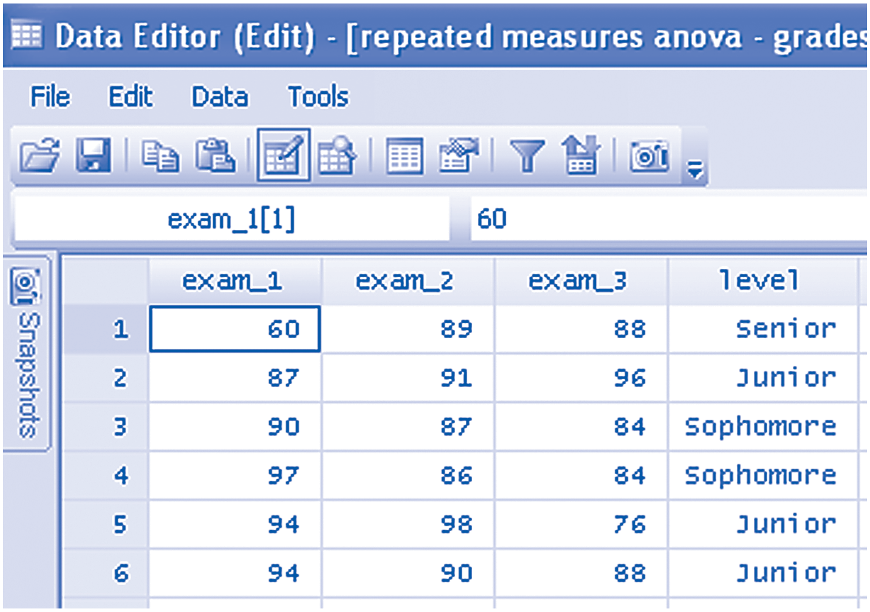
The degrees of freedom, which you will report, come from the *df* column in the table just presented. For example, the *F* test for the full model would be reported as $F(5, 27157) = 573.72$. The first value, 5, comes from the first “Model” row, while the second value, 27157, comes from the “Residual” row.

Our results can be stated in the following way: A factorial ANOVA found a significant difference in income based on both race and gender, $F(5, 27157) = 573.72$, $p < .001$. Specifically, males were found to have significantly higher incomes than females, $F(1, 27157) = 383.95$, $p < .001$. Also, income was found to vary significantly based on race, $F(2, 27157) = 117.93$, $p < .001$. Finally, a significant interaction between race and gender was found, $F(2, 27157) = 55.09$, $p < .001$.

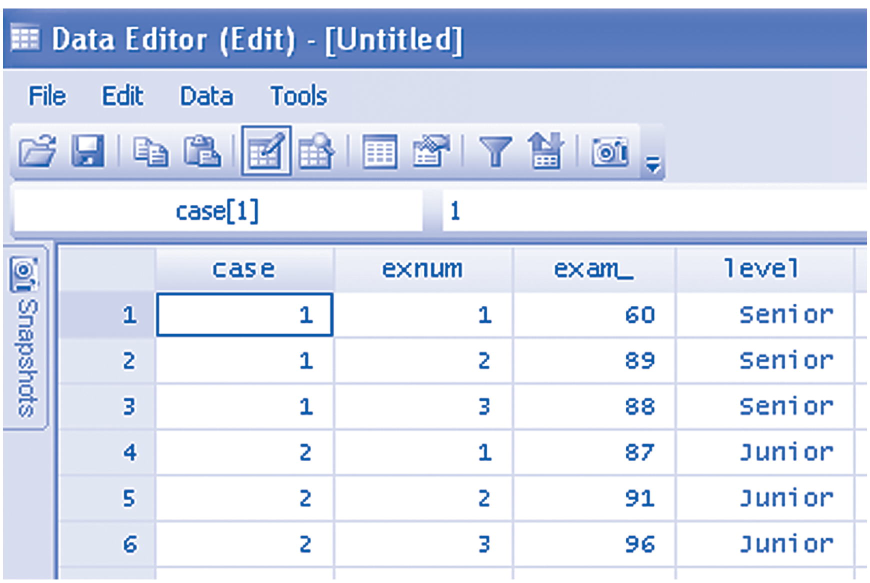
Repeated Measures ANOVA: Stata

Repeated measures ANOVAs are used when your dependent variable consists of a measure that was recorded or measured at several points in time. For example, if you had a set of two or more exam grades for a set of respondents, these data, along with one or more independent predictor variables, could be analyzed using a repeated measures ANOVA. This is the example that I’ll be using in this section. A repeated measures ANOVA could also be used in other situations; for example, if you had a measure for respondents that was taken before and after some medical treatment. Using a repeated measures ANOVA, you can also include predictor variables such as sex and age.

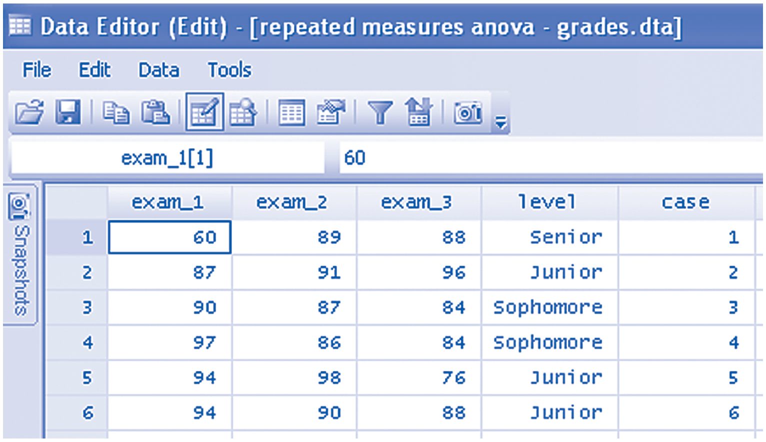
To run a repeated measures ANOVA in Stata, we will first need to “reshape” the data. Currently, the data are in this format:



where each respondent has his or her own single row. We need to get the data into this format:



Where each respondent has three rows, one for each exam score. To do this, we need to first create a new variable to identify respondents by number, such as *case*. Here, I have simply used "1" for the first respondent and have continued from there:



	exam_1	exam_2	exam_3	level	case
1	60	89	88	Senior	1
2	87	91	96	Junior	2
3	90	87	84	Sophomore	3
4	97	86	84	Sophomore	4
5	94	98	76	Junior	5
6	94	90	88	Junior	6

In Stata, you can simply use the command:

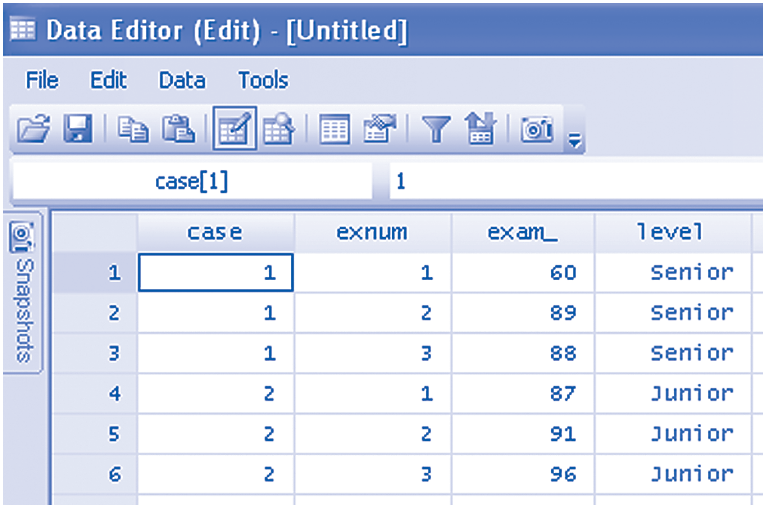
```
gen case=.
```

And then enter the values for *case* by typing ed.
Next, use the following syntax command:

```
reshape long exam_, i(case) j(exnum)
```

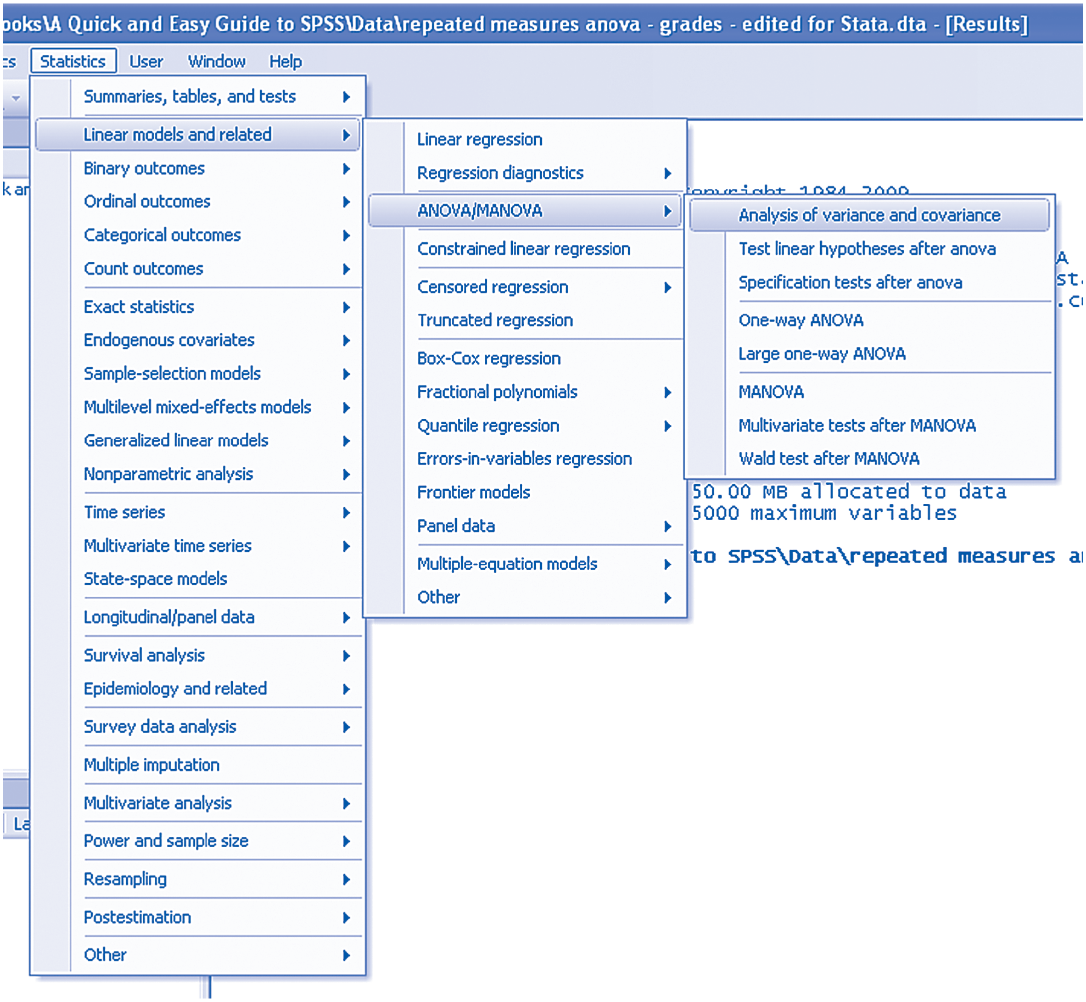
This transforms the data into the necessary "long" format using the *exam* variable. The variable *case* will identify the respondent, and *exnum* will identify the exam number. The new exam variable will be simply *exam_*.

This will transform the data into the necessary format:

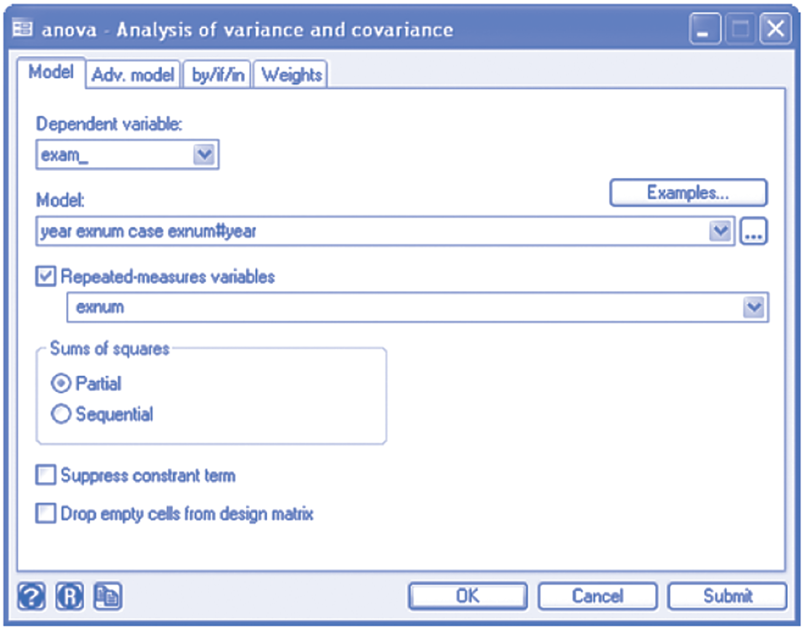


	case	exnum	exam_	level
1	1	1	60	Senior
2	1	2	89	Senior
3	1	3	88	Senior
4	2	1	87	Junior
5	2	2	91	Junior
6	2	3	96	Junior

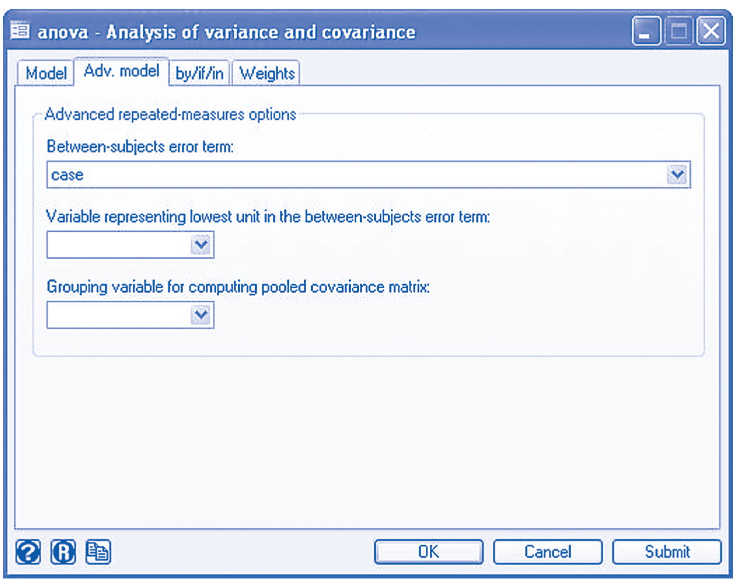
Next, to run a repeated measures ANOVA, make the following menu selection:



This will open the following dialog box:



As you can see, I have specified the new exam grade variable, *exam_*, as the dependent variable in this model. I have included *year* (year at college), *exnum* (exam number), and *case* (respondent number) as independent variables in this model. I have also included *exnum#year*, which is the interaction between exam number and year at college. This tests whether the effect of time (exam number) on exam scores varies significantly by year at college. Also, I have specified that *exnum* is the repeated measures variable. Next, I will click on the *Adv. model* tab:



Here, I will need to specify *case*, the respondent number, as the between-subjects error term. Clicking *OK* results in the following output:

```
. anova exam_ year exnum case exnum#year, repeated(exnum) bse(case)

                           Number of obs =       111     R-squared     =  0.5646
                           Root MSE      =  13.1472     Adj R-squared =  0.2743

                  Source | Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model | 14790.4535    44  336.146671        1.94     0.0070
                         |
                    year | 1608.91667     3  536.305556        3.10     0.0325
                   exnum | 918.972718     2  459.486359        2.66     0.0776
                    case | 9418.74286    33   285.41645        1.65     0.0421
              exnum#year | 509.048134     6  84.8413556        0.49     0.8129
                         |
                Residual | 11407.9429    66  172.847619
              -----------+----------------------------------------------------
                   Total | 26198.3964   110   238.16724

Between-subjects error term:  case
                     Levels:  37        (33 df)
     Lowest b.s.e. variable:  case

Repeated variable: exnum
                                         Huynh-Feldt epsilon        =  0.8152
                                         Greenhouse-Geisser epsilon =  0.7223
                                         Box's conservative epsilon =  0.5000

                                                 ------------ Prob > F ------------
                  Source |     df      F    Regular    H-F      G-G      Box
              -----------+----------------------------------------------------
                   exnum |      2    2.66   0.0776   0.0895   0.0960   0.1125
              exnum#year |      6    0.49   0.8129   0.7778   0.7566   0.6910
                Residual |     66
              -----------+----------------------------------------------------
```

Here, the model was found to be significant, having an F statistic of 1.94, with $p < .01$. Year at college was found to be a significant predictor of exam scores, having an F statistic of 3.10, with $p < .05$. Next, the exam number was not found to have a significant effect on exam scores. The interaction between exam number and year at college was also not found to be significant at the $p < .05$ level.

The syntax command, shown in the output, is as follows:

```
anova exam_ year exnum case exnum#year,
      repeated(exnum) bse(case)
```

SECTION 4: SUMMARY

This chapter covered Pearson's *r*, chi-square, the *t*-test, and the ANOVA. Pearson's *r*, a correlation coefficient, is used to determine the strength and direction of the relationship between two continuous variables. Chi-square is used to show whether or not there is a relationship between two categorical variables. It can also be used to test whether or not a number of outcomes are occurring in equal frequencies or not, or conform to a certain distribution. Both the *t*-test and the ANOVA are used to test differences in scores between groups. While the *t*-test can only be used to test the differences between two groups on some continuous variable, the ANOVA can be used to test the differences between two or more groups on a continuous variable. When conducting an ANOVA on more than two groups, it is necessary to select a post hoc comparison test in order to determine between which specific groups there is a significant difference. A one-way ANOVA includes only one independent, predictor variable, while factorial ANOVAs include two or more. Also, repeated measures ANOVAs are used to look at a dependent variable that is measured at multiple points in time. The next chapter will cover linear regression, which is a particular form of regression that is used when your dependent variable is continuous. Regression is a powerful statistical tool as it allows you to determine the effect of one independent variable on your dependent variable while holding any number of other independent variables constant. Starting with the following chapter, we will begin constructing and analyzing models that include more than one independent variable, moving on from bivariate (two variables) statistics and beginning our journey into what is called multivariate statistics.

RESOURCES

You can find more information about IBM SPSS and how to purchase it by navigating to the following Web site: www.spss.com/software/statistics/

You can find more information about Stata and how to purchase it by navigating to the following Web site: www.stata.com

This book's Web site can be found at the following location: www.sagepub.com/kremelstudy