

Running a One-Way ANOVA in SPSS 21-22

This guide uses the *RandStudy* data set available on Blackboard.

This example investigates the question of whether a person who has a positive view of making money from crime (i.e., that he thinks he will make money from criminal activities) commits more crimes than a person who does not have such an optimistic view. In other words, do people who think they will succeed financially at crime actually DO MORE crimes than other not-so-optimistic people?

The Rand Study data set is the result of a famous research project carried out by the Rand Corporation in the 1970s in a California prison. Researchers went in and asked 410 prisoners about their level of criminal activity in the three years before the arrest for which they were sent to prison. The researchers were totally shocked by the information they were given (and believed every word). The term “criminal career” was born, and “get tough on crime” era had begun. The data, as you will see when you look at the variables, includes many variables that express attitudes about crime and an honest lifestyle. Each of these attitude variables has 5 levels – they are ordinal. In this example, we will look at the average number of crimes committed by the prisoners, grouped by their attitudes. We will use ANOVA (Analysis of Variance) because there are more than two groups to be compared.

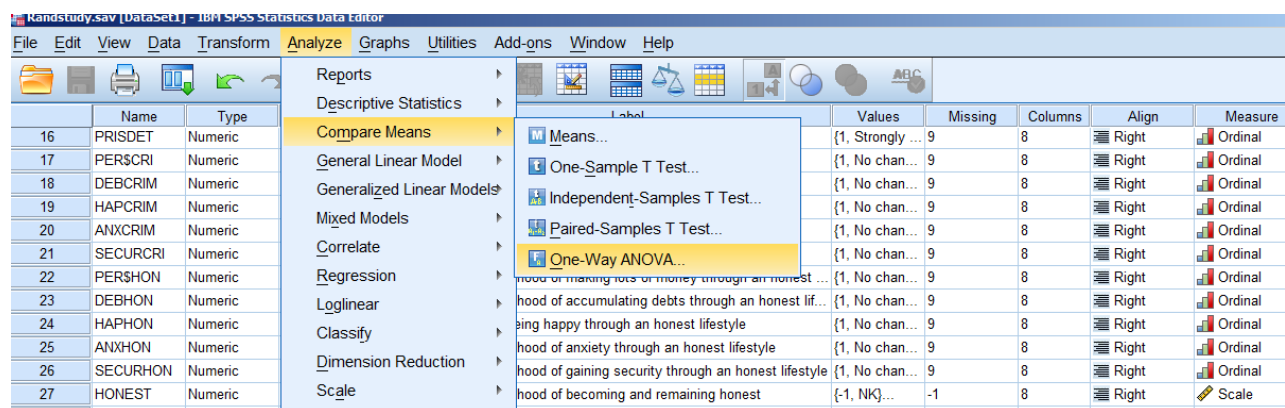
Before you begin, look at the frequencies for the two variables:

- Independent variable = **PER\$CRI** (Perceived likelihood of making lots of money through crime)
- Dependent variable = **CONV#1** (the number of convictions in the 3-year period)

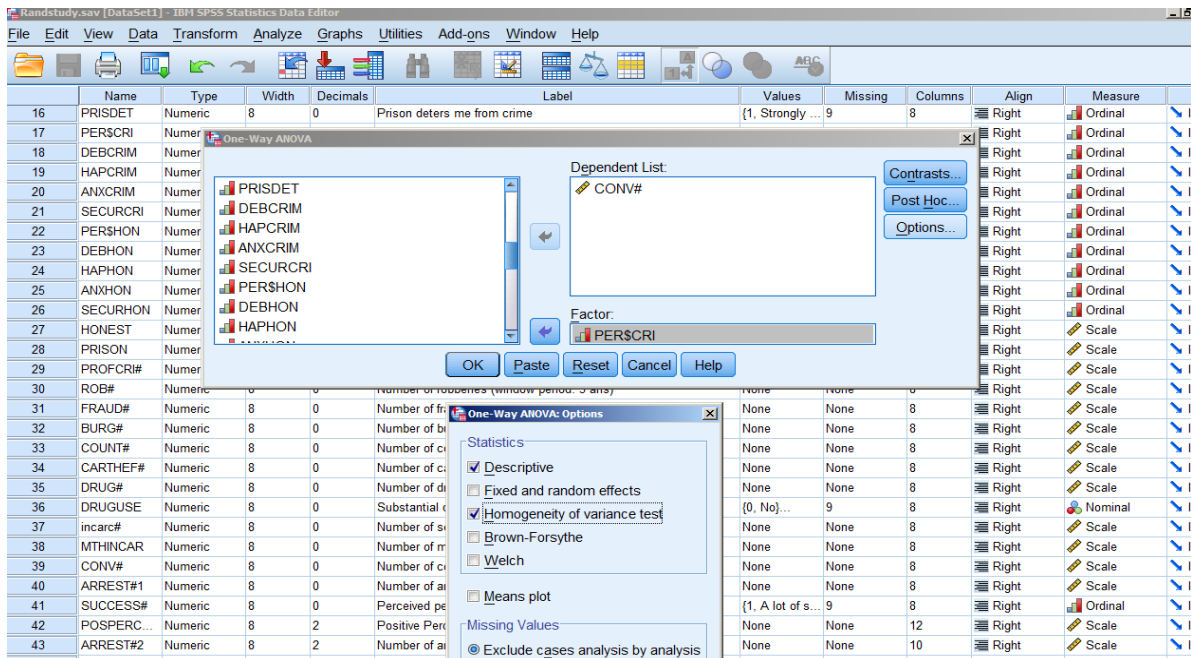
We are going to use a one-way ANOVA to test the **null hypothesis** that there is no difference in the average number of convictions between the different groups of prisoners defined by their attitudes toward the results of crime (or that any difference seen in our sample is simply due to sampling error). This hypothesis applies to all who engage in criminal behavior – we test the hypothesis with our sample.

Written in notation, the Null Hypothesis would be: $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

To run a one-way ANOVA, go to Analyze → Compare Means → One-Way ANOVA.



Select One-Way ANOVA and place the variables in the window shown below.



The Dependent variable is CONV# - but you could have explored more than one dependent variable.

The Independent variable, PER\$CRI, is called a FACTOR in ANOVA. It is the “grouping variable.”

From the OPTIONS button, open the Options window shown below the variables window. Select two options: **Descriptive** and **Homogeneity of Variance Test**. First, we request descriptive statistics, so that if we find a statistically significant difference somewhere, we can describe the nature of the difference (which group or groups is higher on the dependent variable than the others, etc.). We also want to request homogeneity of variance tests – unlike in the independent- samples t test, SPSS will not automatically run this test so we need to request it separately. We can get both of these things by clicking on the Options button.

Descriptives

Number of convictions (window period: 3 yrs)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					No chance	33		
Small chance	81	1.93	2.649	.294	1.34	2.51	0	16
Average chance	131	1.76	1.814	.158	1.45	2.08	0	9
Strong chance	99	2.10	2.443	.246	1.61	2.59	0	17
Certain	61	1.87	2.085	.267	1.33	2.40	0	11
Total	405	1.87	2.175	.108	1.66	2.08	0	17

As you can see, the Descriptives result table shows the information about each group – the number of prisoners in each group (N), the average number of convictions (Mean) and the variability in each group (Std. Deviation, Minimum and Maximum). We see that the averages have a small range, from 1.48 to 2.10, and that the variability is large in each group (at least 0 to 5 and as much as 0 to 17).

The ANOVA statistic is an index or ratio of the difference between group means (numerator) divided by the overall variability (denominator). In this example, the observed difference in the means is not large, and the overall variability (the expected) is large. There is more to the math of ANOVA than this ... consult the text for more detail ... but this ratio is the heart of the F Statistic.

Are the groups equal (homogeneous) in variance? Yes ... according the result of the Levene's test.

Test of Homogeneity of Variances

Number of convictions (window period: 3 yrs)

Levene Statistic	df1	df2	Sig.
.855	4	400	.491

The Levene Statistic is small, less than 1855.

The probability of finding the observed variability by chance is high491 – far above the normal .05 criterion. We accept the Null Hypothesis.

Conclusion: The variances are equivalent in the grouping defined by the factor, PER\$CRI.

The ANOVA result

ANOVA

Number of convictions (window period: 3 yrs)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.921	4	2.980	.628	.643
Within Groups	1899.403	400	4.749		
Total	1911.323	404			

The variability between the group means (Sum of Squares for Between Groups / DF = 4) is 2.980.

The variability within the groups (Sum of Squares for Within Groups / df = 404) = 4.749.

Mean square for Between Groups / mean square for Within Groups produces the F value: .628

RESULT: As before: there are two ways to understand whether this result is a consequence of random variation or whether the attitudes that created the groups influence behavior: the Test Statistic or F value and the statistical significance or sig. The F statistic is small and the probability of finding the differences that we see in the Descriptives table is large = **we accept the Null Hypothesis.**

To find an example of an ANOVA result that would lead us to REJECT the Null Hypothesis, I had to open another dataset. In the NELS dataset, the output below shows the result of this research question:

Does the area (environment) student lives in (Rural, Suburban and Urban) influence math achievement in the 12th grade?

The variables are: Dependent = **achmat12** (Math Achievement in the 12th Grade – a test score)
 Independent = **urban** (Urbanicity – nominal with values = rural, suburban, urban)

Descriptives

Math Achievement in Twelfth Grade

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					Urban	123		
Suburban	215	57.5307	7.74979	.52853	56.4889	58.5725	34.36	71.05
Rural	162	55.0714	7.64271	.60047	53.8855	56.2572	35.56	71.12
Total	500	56.9066	7.88403	.35258	56.2139	57.5994	34.36	71.12

The Descriptives table does show differences in the mean scores for children in these different areas. And the difference between Rural children at 55.07 and other children at 57.53 or 58.23 is more than two standard errors ... remember the z diagrams? If the sample means is more than two standard errors away from the mean ... it is in or near the critical zone ...

Test of Homogeneity of Variances

Math Achievement in Twelfth Grade

Levene Statistic	df1	df2	Sig.
.482	2	497	.618

No issues with Homogeneity of Variance ... groups have equal variance ...

ANOVA

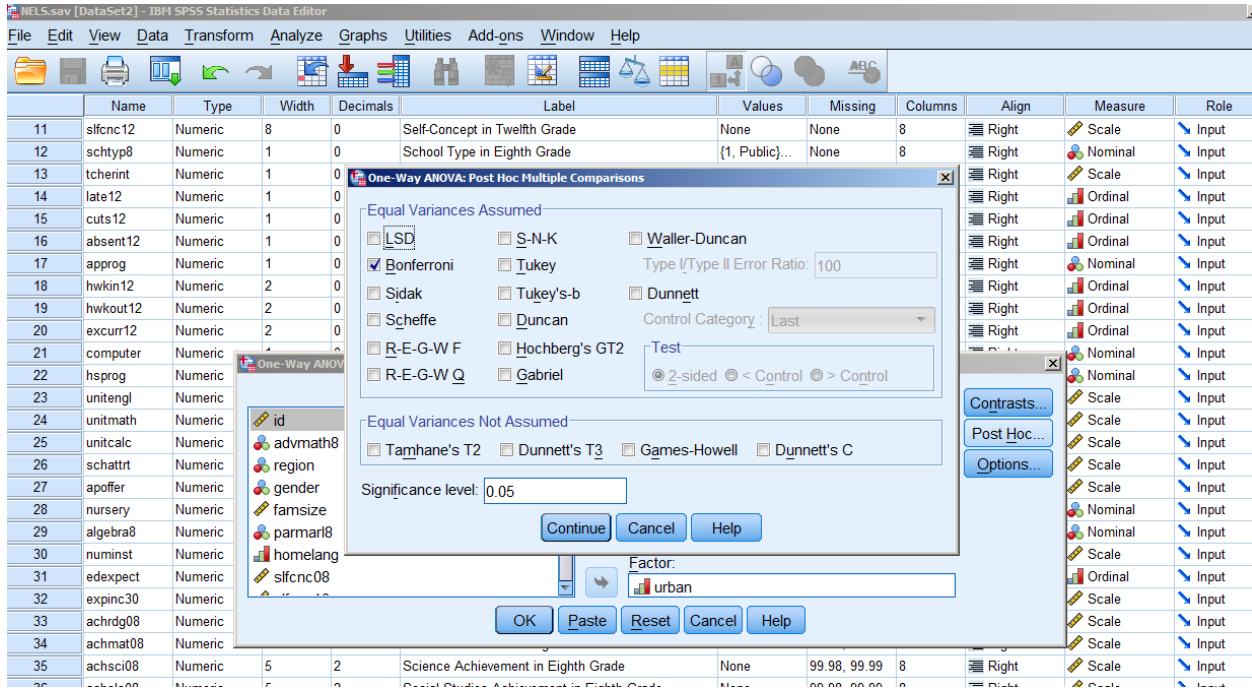
Math Achievement in Twelfth Grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	845.751	2	422.875	6.966	.001
Within Groups	30171.032	497	60.706		
Total	31016.783	499			

RESULT: The F value is larger at F= 6.966 and the probability of finding this result by chance is one in 1,000 or P = .001. This time we conclude that the independent variable, urbanicity, DOES influence the level of math achievement of students in these areas. One more step see next page.

When an ANOVA procedure generates a result that leads to the REJECTION of the Null Hypothesis, there is one more step with the software. This step, called a POST HOC test, is only taken when the result leads you to reject the Null.

If you go back to the ANOVA screen, this time open the POST HOC window: If the Levene's Test result leads you to accept the Null (Equal Variances Assumed), the use the Bonferroni test. (Otherwise you would choose from the tests that do not require equal variance).



The resulting POST HOC table will allow you to see exactly what contrasts (group comparisons) are statistically significant – and provides an asterisk to make you see it!

Multiple Comparisons

Dependent Variable: Math Achievement in Twelfth Grade

Bonferroni

(I) Urbanicity	(J) Urbanicity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Urban	Suburban	.70223	.88085	1.000	-1.4137	2.8181
	Rural	3.16157*	.93181	.002	.9232	5.3999
Suburban	Urban	-.70223	.88085	1.000	-2.8181	1.4137
	Rural	2.45934*	.81061	.008	.5122	4.4065
Rural	Urban	-3.16157*	.93181	.002	-5.3999	-.9232
	Suburban	-2.45934*	.81061	.008	-4.4065	-.5122

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