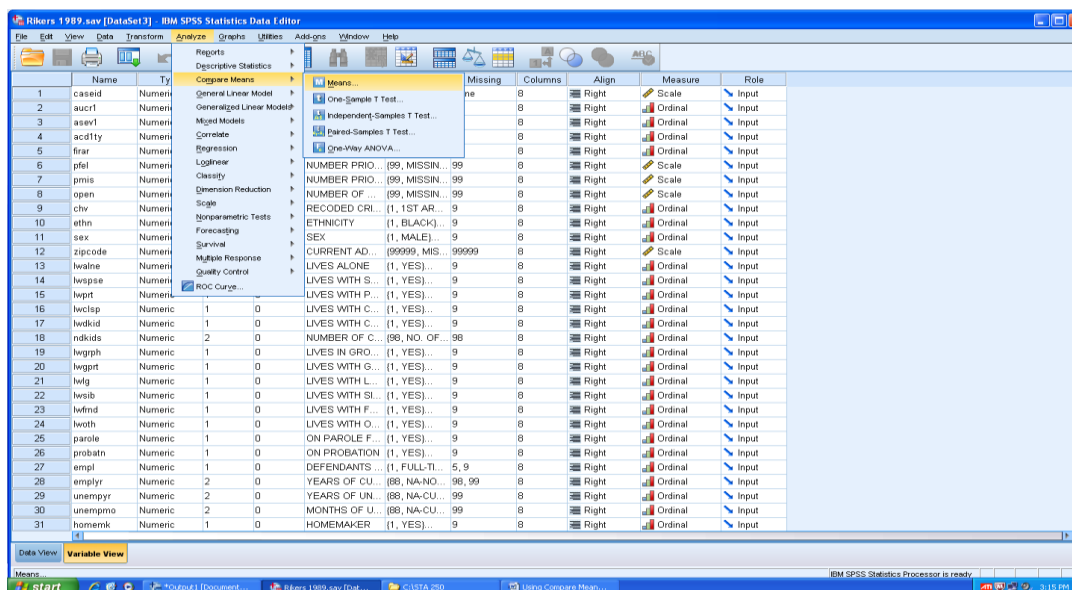**Using Compare Means in IBM SPSS versions 21-22 to look at the comparative averages on a variable of interest (a dependent variable) for several subgroups of a second variable (an independent variable).**

Very often we are interested in comparing groups. Do students with children do better than students who do not have children? Does a new program for freshmen help them pass the Compass test – compared to those who did not have the benefit of the program? Does locking up juveniles help them stay arrest-free longer than putting them on probation?
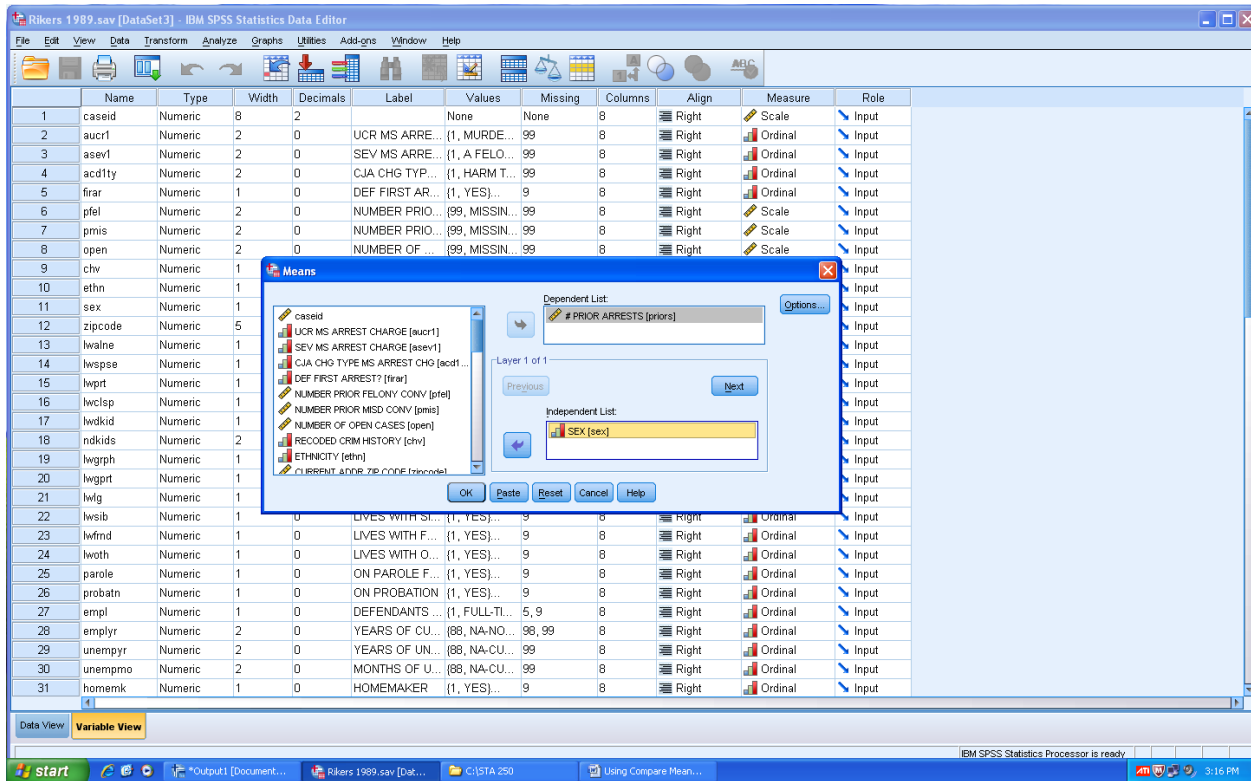
The data sets for all three questions would be record of individuals – students in the first one, freshmen in the second one and juvenile offenders in the third one. All of these questions are similar in structure: the groups are categories of some nominal variable (we call this the independent variable because we think it influences the second variable) and the variable of interest is a count – something that can be measured at the ratio level.

- The variable Student Status has two categories > 1 = Has children, 2= No children. The variable we would measure to compare these groups would be GPA – a ratio-level variable. We would look at the average GPA for students with children and the average GPA for students who do not have children.
- The freshmen example is a file with all incoming students for some semester. The groups would be defined by a Program variable > 1 = student who took the program; 2= students who did not have the program. And we would look at the average scores on the Compass test.
- The juvenile offender example is similar. We would look at the number of days free of arrest for two groups > 1 = Jail; 2= probation. Again, we would be looking at the average days free of arrest for the juveniles in each group.

Compare Means provides a group of procedures, of which **MEANS** is the simplest one.

In the screen below, **sex** is being used to define the groups and (the average number of prior arrests) **priors** is the variable of interest – the Dependent Variable.

The results show that for the 1030 cases for which there was data for both variables, on average the female arrestees have more prior arrests than the male arrestees.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
# PRIOR ARRESTS * SEX	1030	99.2%	8	.8%	1038	100.0%

Report

PRIOR ARRESTS

SEX	Mean	N	Std. Deviation
MALE	5.06	915	7.500
FEMALE	5.74	115	11.103
Total	5.14	1030	7.979