Assumption of Homoscedasticity

Homoscedasticity
(also referred to as homogeneity of variance)
(also referred to as uniformity of variance)

Transformations

Assumption of normality script

Practice problems

Assumption of Homoscedasticity

- Homoscedasticity refers to the assumption that that the dependent variable exhibits similar amounts of variance across the range of values for an independent variable.

- While it applies to independent variables at all three measurement levels, the methods that we will use to evaluation homoscedasticity requires that the independent variable be non-metric (nominal or ordinal) and the dependent variable be metric (ordinal or interval). When both variables are metric, the assumption is evaluated as part of the residual analysis in multiple regression.

Evaluating homoscedasticity

- Homoscedasticity is evaluated for pairs of variables.
- There are both graphical and statistical methods for evaluating homoscedasticity .
- The graphical method is called a boxplot.
- The statistical method is the Levene statistic which SPSS computes for the test of homogeneity of variances.
- Neither of the methods is absolutely definitive.

Transformations

- When the assumption of homoscedasticity is not supported, we can transform the dependent variable variable and test it for homoscedasticity . If the transformed variable demonstrates homoscedasticity, we can substitute it in our analysis.
- We use the sample three common transformations that we used for normality: the logarithmic transformation, the square root transformation, and the inverse transformation.
- All of these change the measuring scale on the horizontal axis of a histogram to produce a transformed variable that is mathematically equivalent to the original variable.

When transformations do not work

- When none of the transformations results in homoscedasticity for the variables in the relationship, including that variable in the analysis will reduce our effectiveness at identifying statistical relationships, i.e. we lose power.

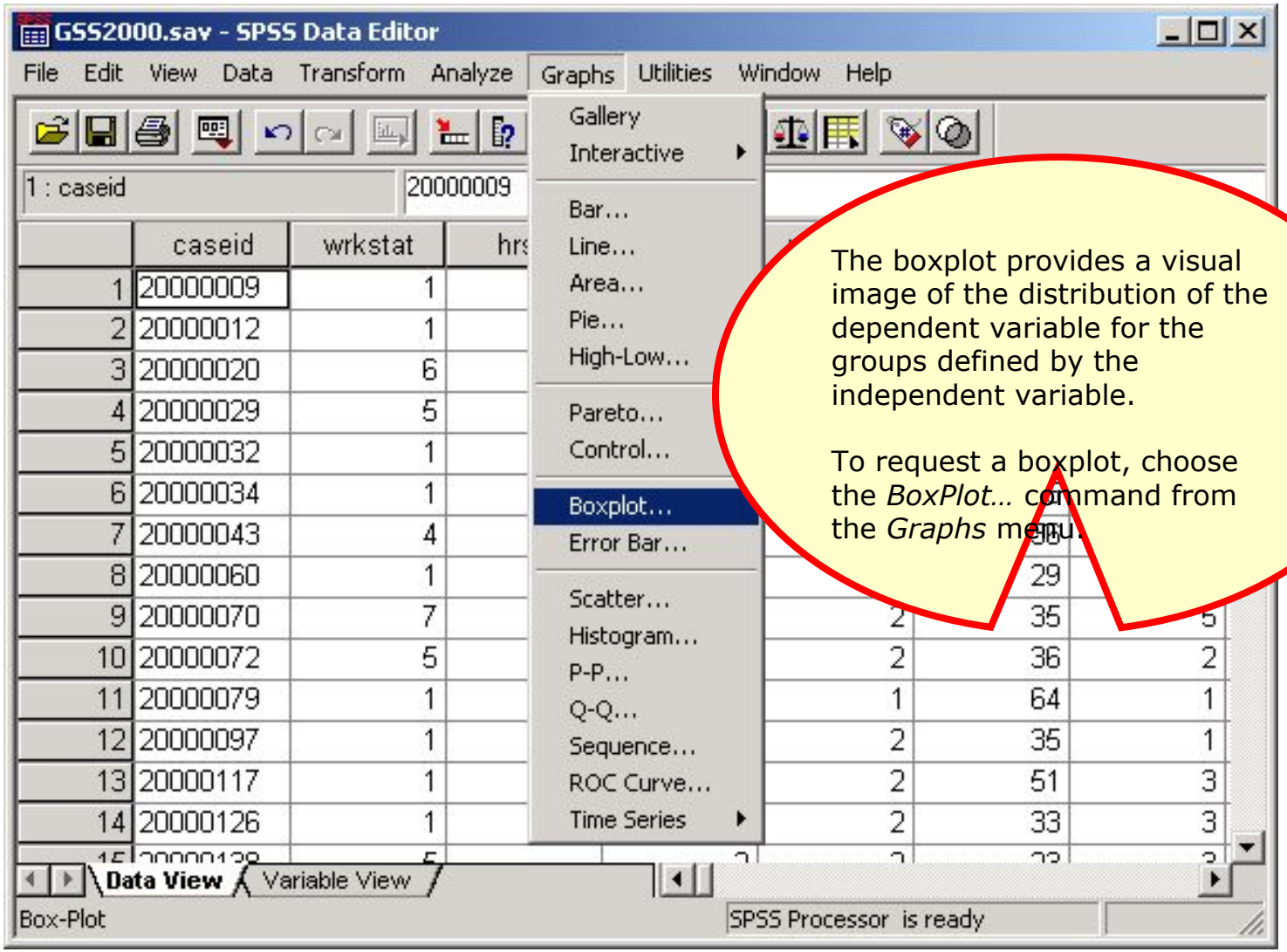
Problem 1

In the dataset GSS2000.sav, is the following statement true, false, or an incorrect application of a statistic? Use 0.01 as the level of significance.

Based on a diagnostic hypothesis test for homogeneity of variance, the variance in "highest academic degree" is homogeneous for the categories of "marital status.“

1. True
2. True with caution
3. False
4. Incorrect application of a statistic

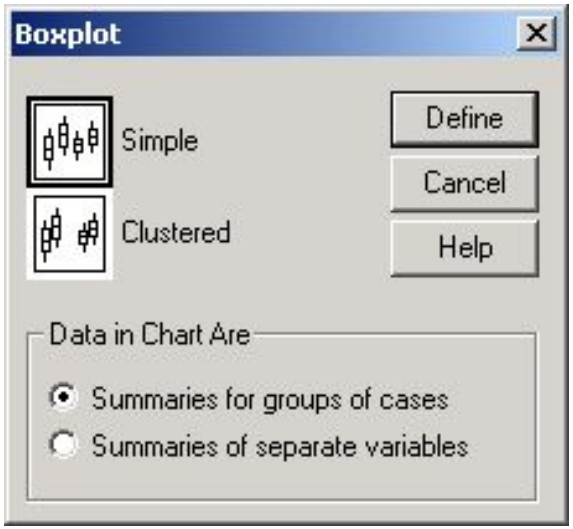
Request a boxplot



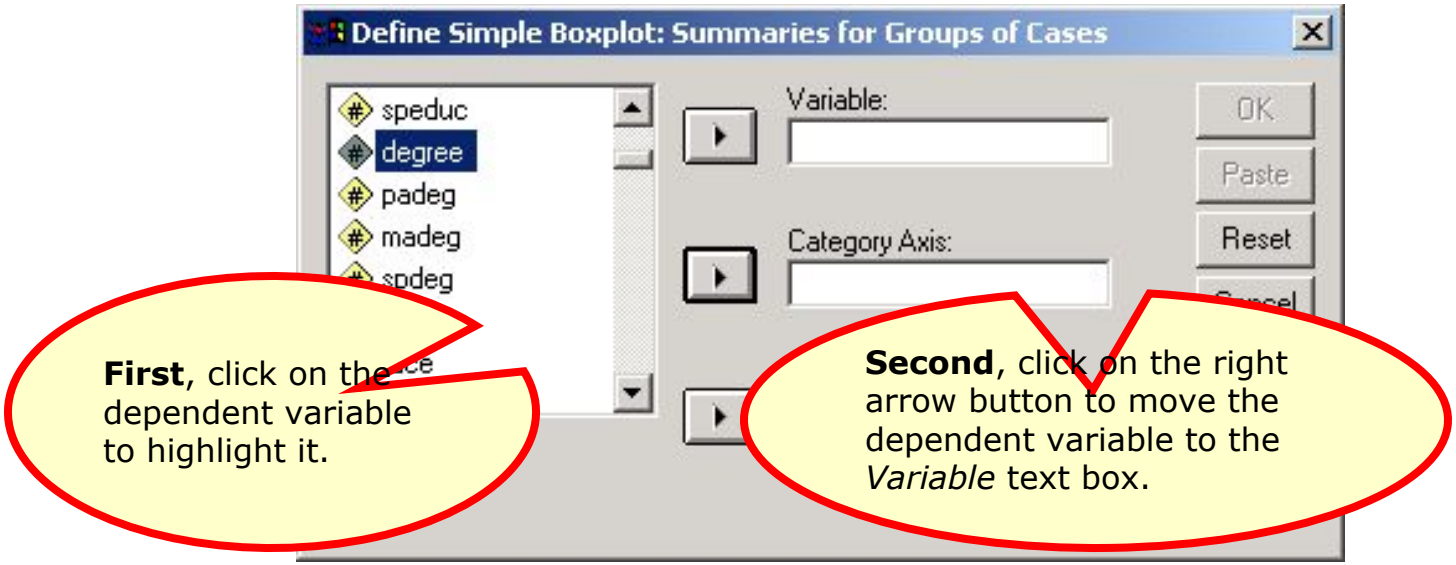
Specify the type of boxplot

First, click on the *Simple* style of boxplot to highlight it with a rectangle around the thumbnail drawing.

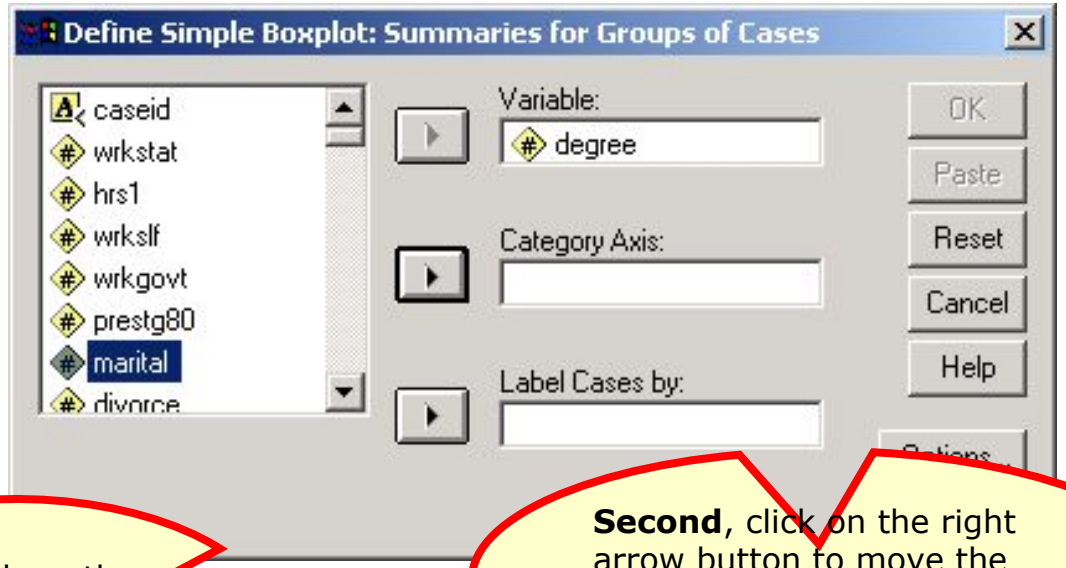
Second, click on the *Define* button to specify the variables to be plotted.



Specify the dependent variable



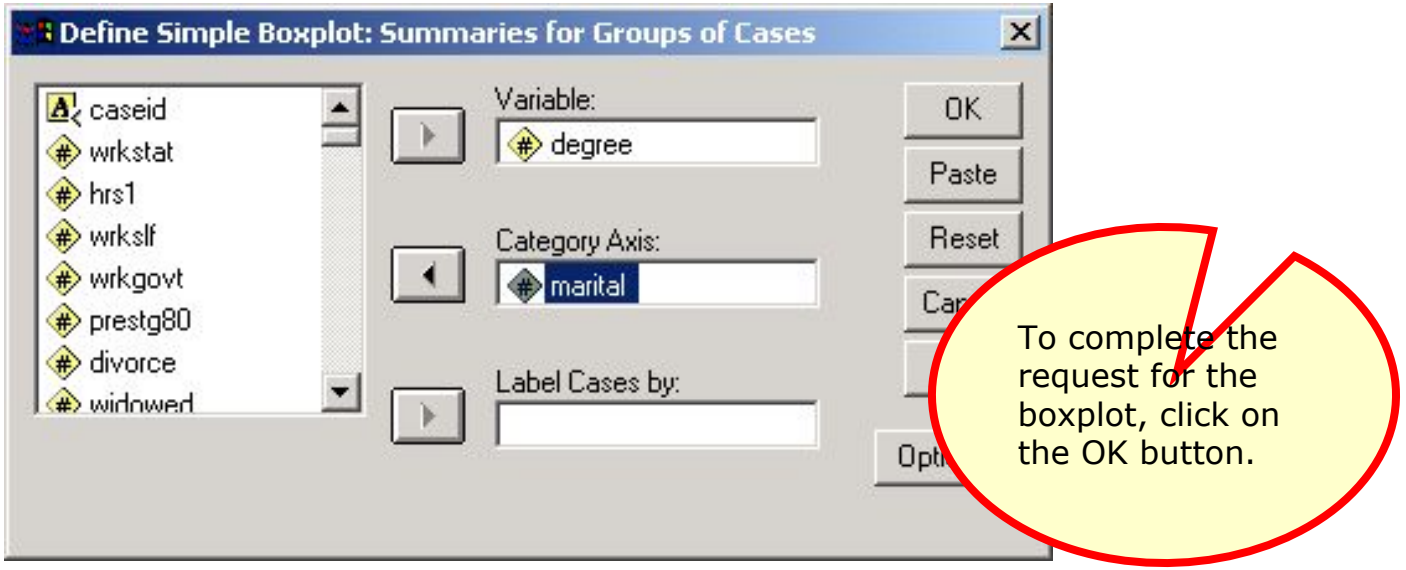
Specify the independent variable



First, click on the independent variable to highlight it.

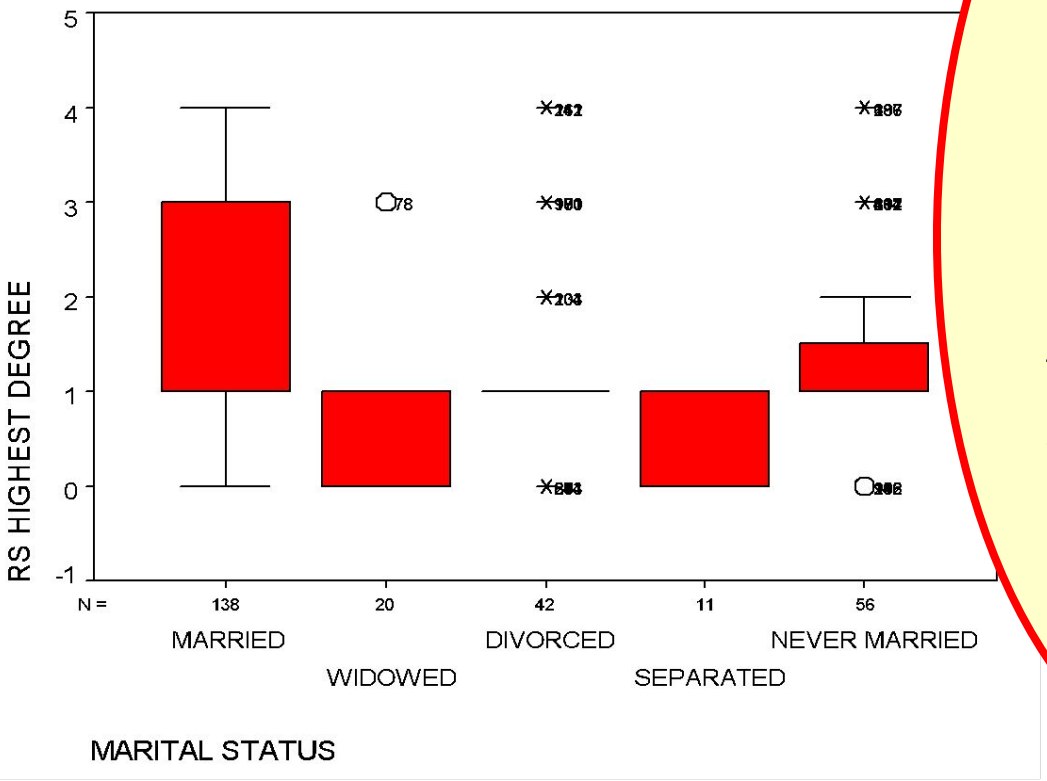
Second, click on the right arrow button to move the independent variable to the *Category Axis* text box.

Complete the request for the boxplot



The boxplot

Each red box shows the middle 50% of the cases for the group, indicating how spread out the group of scores is.

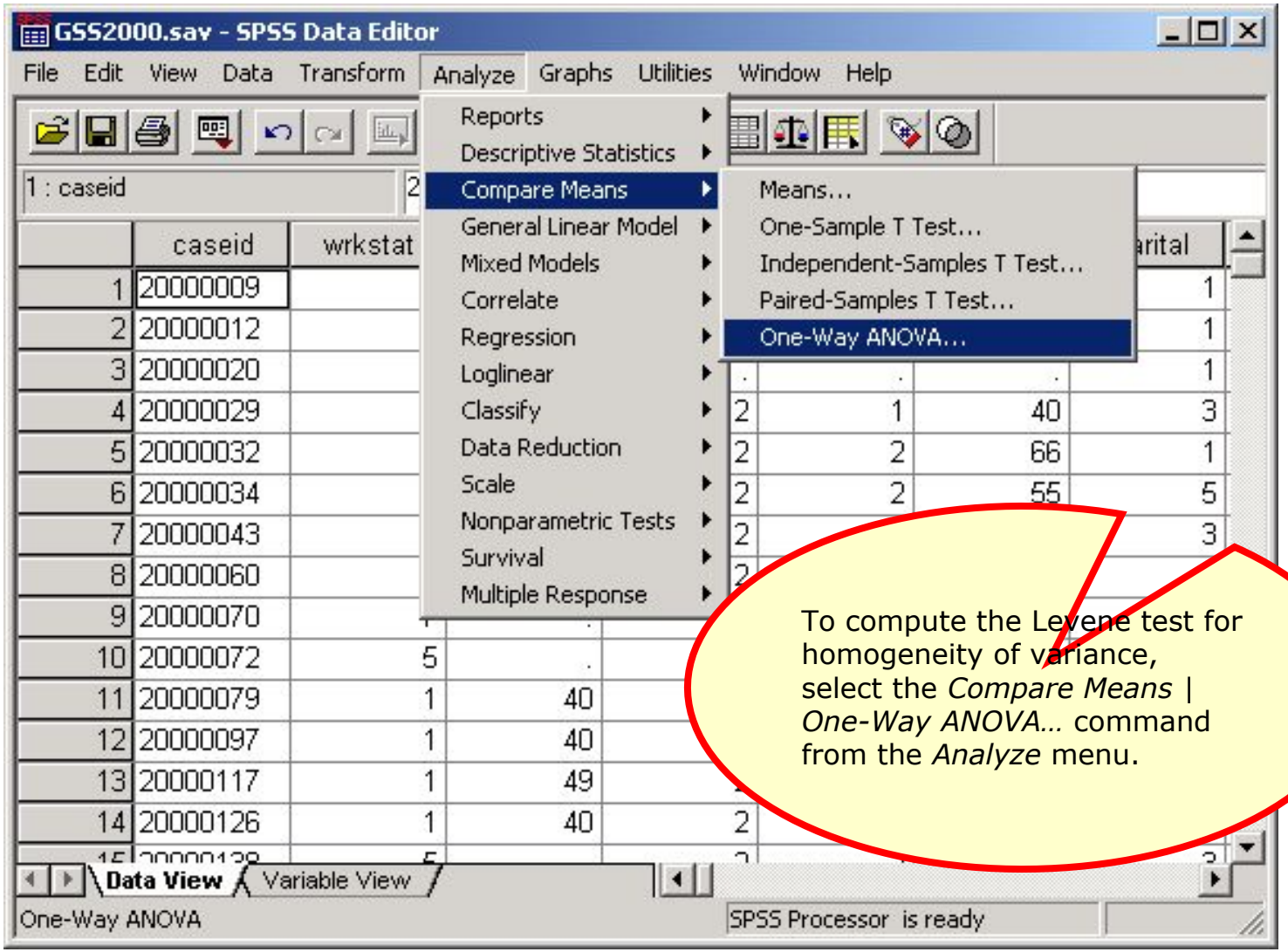


If the variance across the groups is equal, the height of the red boxes will be similar across the groups.

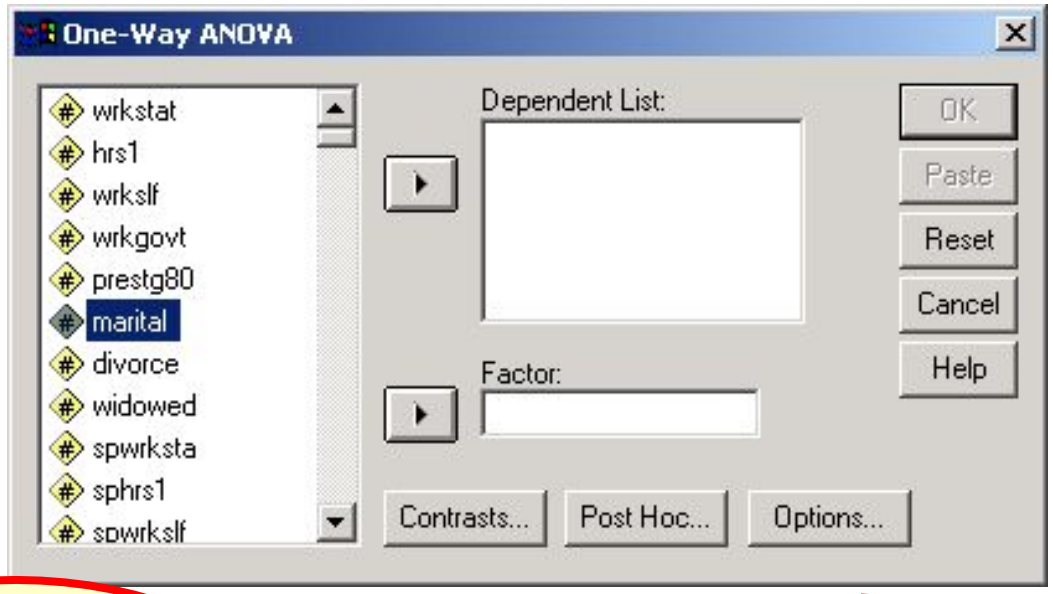
If the heights of the red boxes are different, the plot suggests that the variance across groups is not homogeneous.

The married group is more spread out than the other groups, suggesting unequal variance.

Request the test for homogeneity of variance



Specify the independent variable



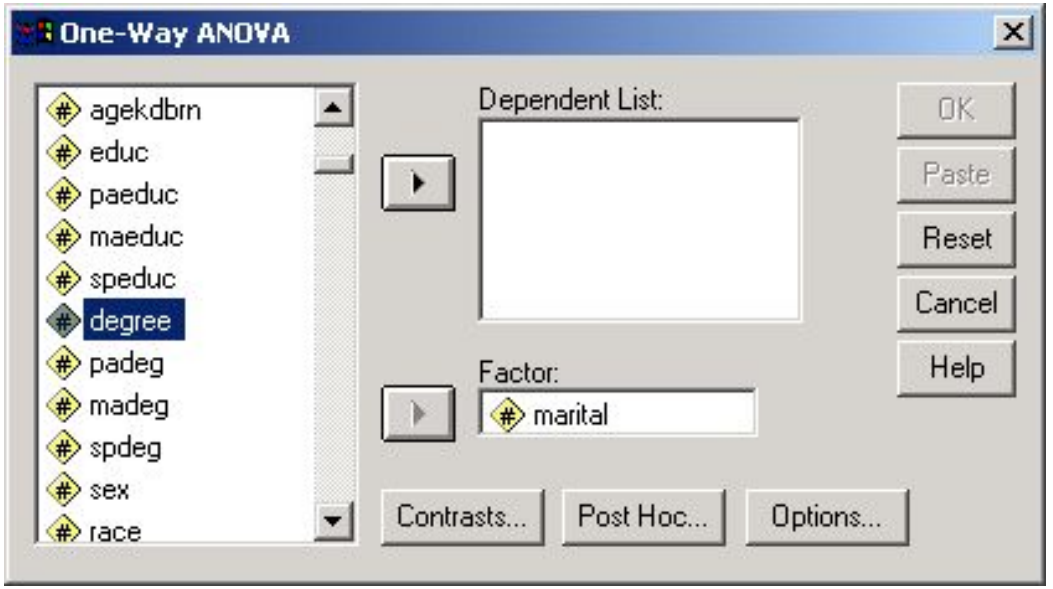
First, click on the independent variable to highlight it.

Second, click on the right arrow button to move the independent variable to the *Factor* text box.

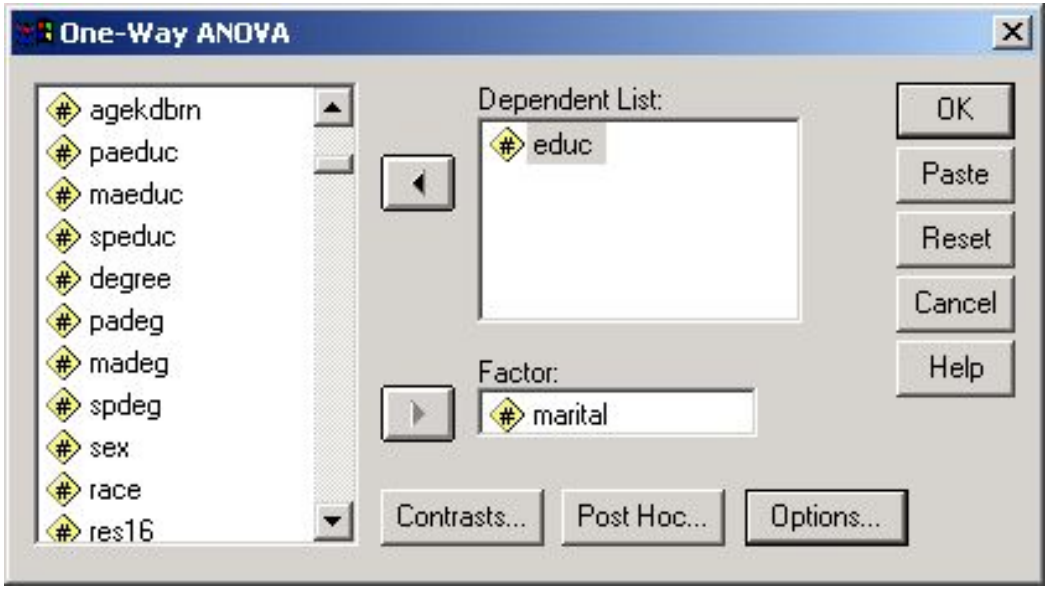
Specify the dependent variable

First, click on the dependent variable to highlight it.

Second, click on the right arrow button to move the dependent variable to the *Dependent List* text box.

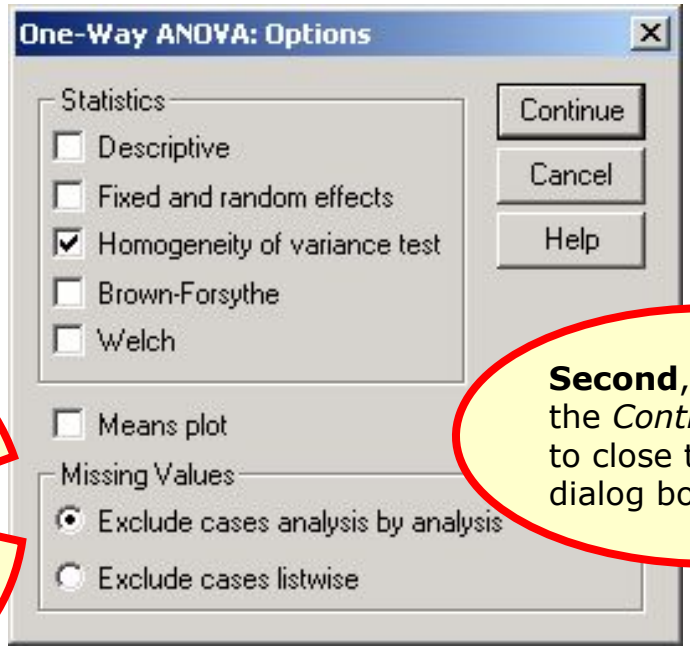


The homogeneity of variance test is an option



Click on the Options… button to open the options dialog box.

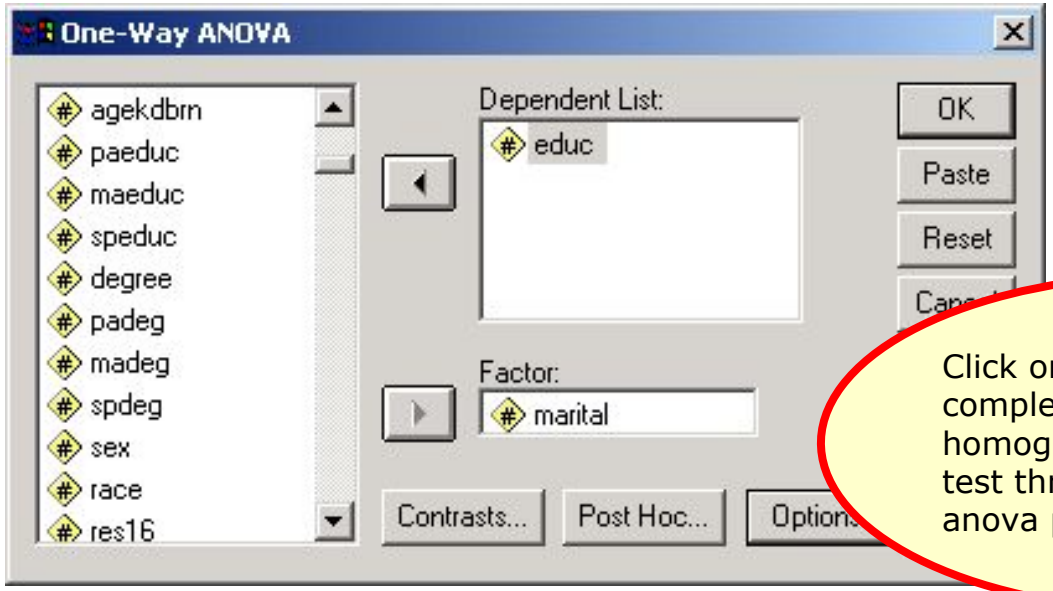
Specify the homogeneity of variance test



First, mark the checkbox for the *Homogeneity of variance test*. All of the other checkboxes can be cleared.

Second, click on the *Continue* button to close the options dialog box.

Complete the request for output



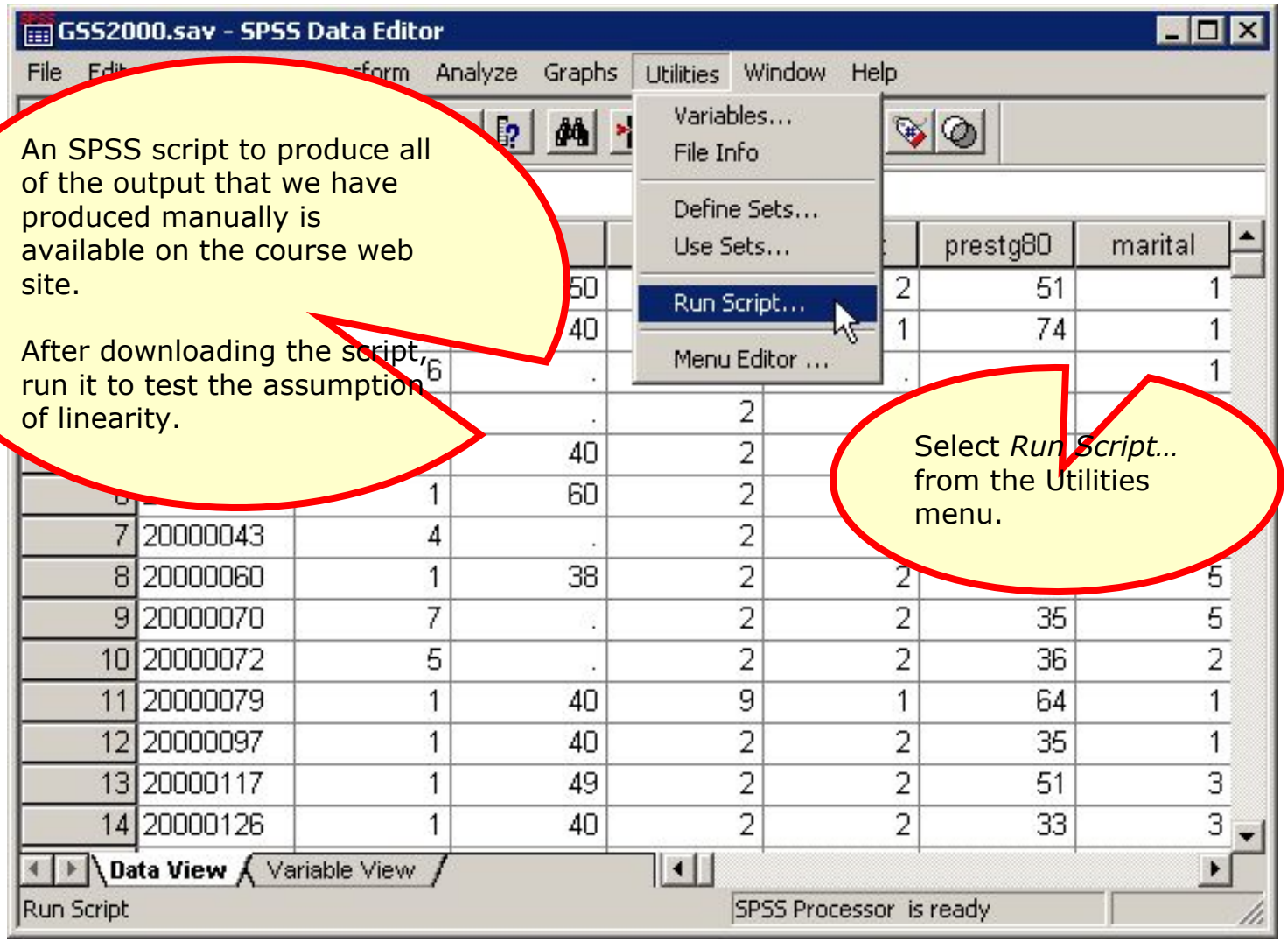
Interpreting the homogeneity of variance test

The null hypothesis for the test of homogeneity of variance states that the variance of the dependent variable is equal across groups defined by the independent variable, i.e., the variance is homogeneous.

Since the probability associated with the Levene Statistic (<0.001) is less than or equal to the level of significance, we reject the null hypothesis and conclude that the variance is not homogeneous.

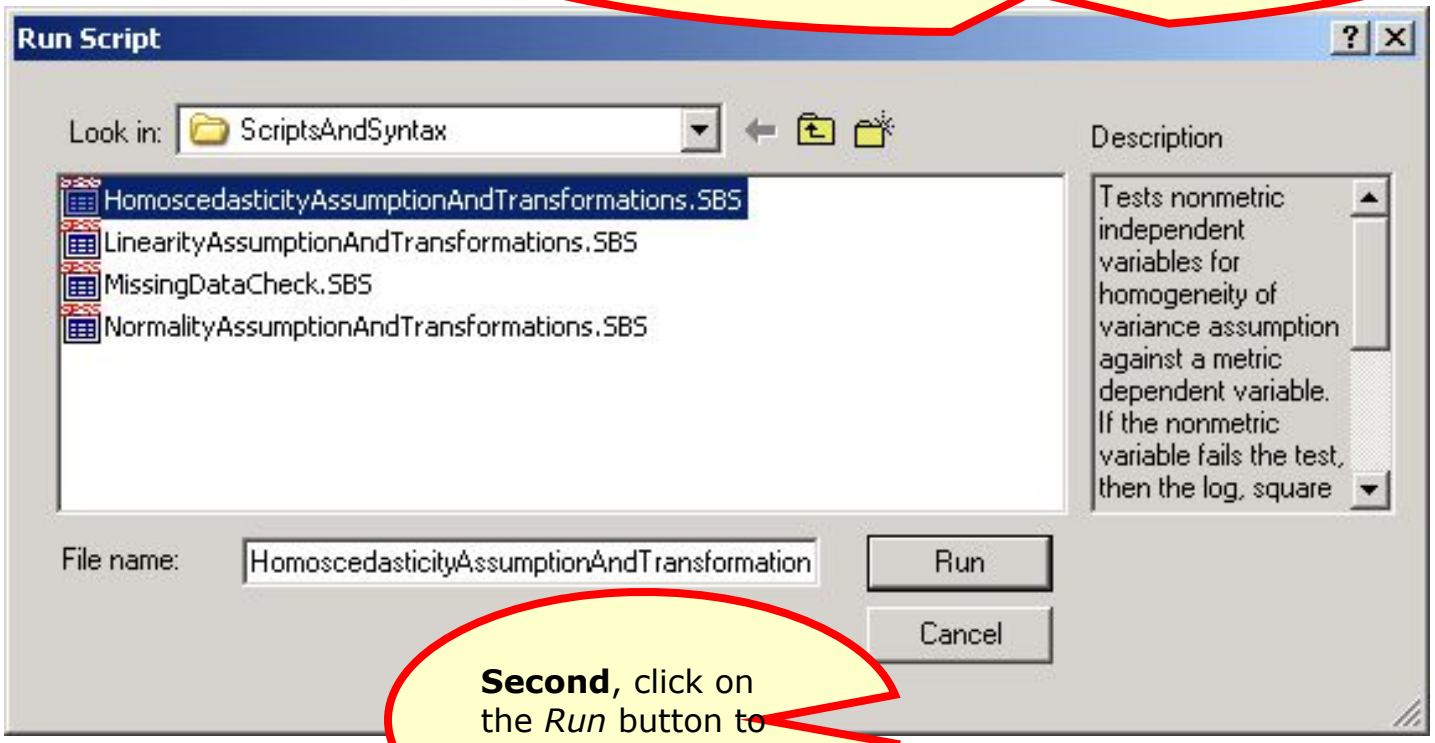
The answer to the question is **false**.

The assumption of homoscedasticity script



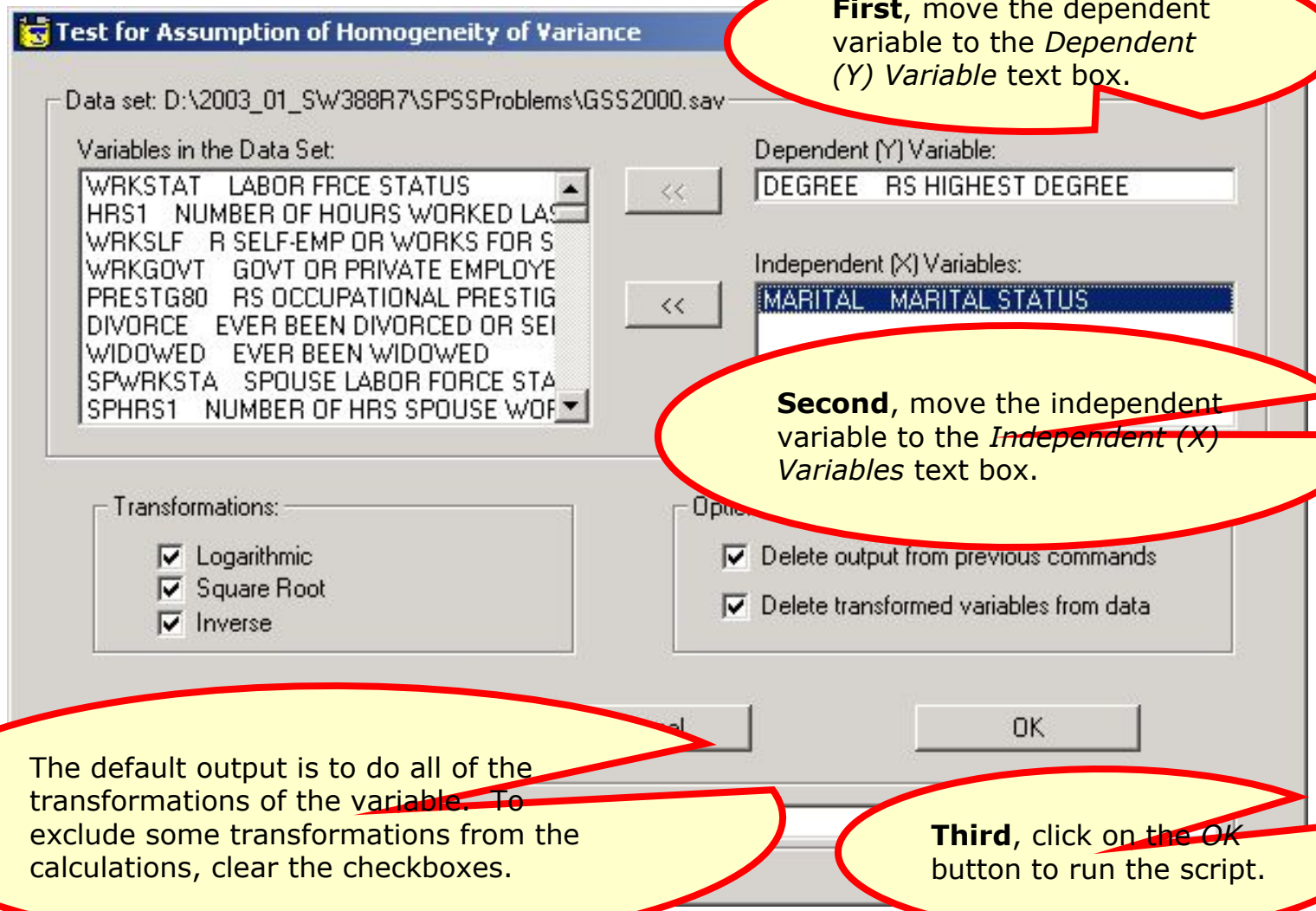
Selecting the assumption of homoscedasticity script

First, navigate to the folder containing your scripts and highlight the script: HomoscedasticityAssumptionAndTransformations.SBS



Second, click on the *Run* button to activate the script.

Specifications for homoscedasticity script



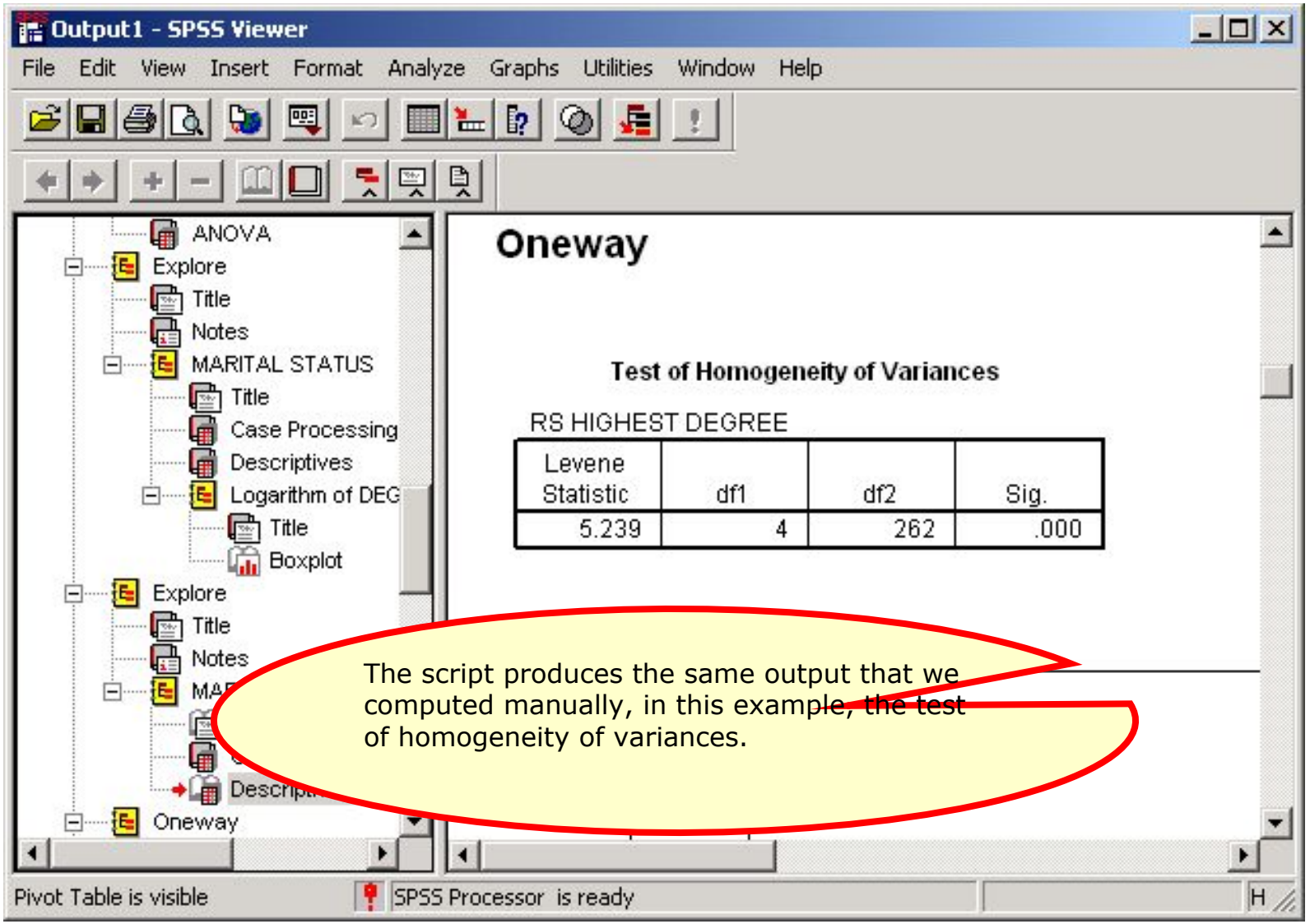
First, move the dependent variable to the *Dependent (Y) Variable* text box.

Second, move the independent variable to the *Independent (X) Variables* text box.

The default output is to do all of the transformations of the variable. To exclude some transformations from the calculations, clear the checkboxes.

Third, click on the *OK* button to run the script.

The test of homogeneity of variance



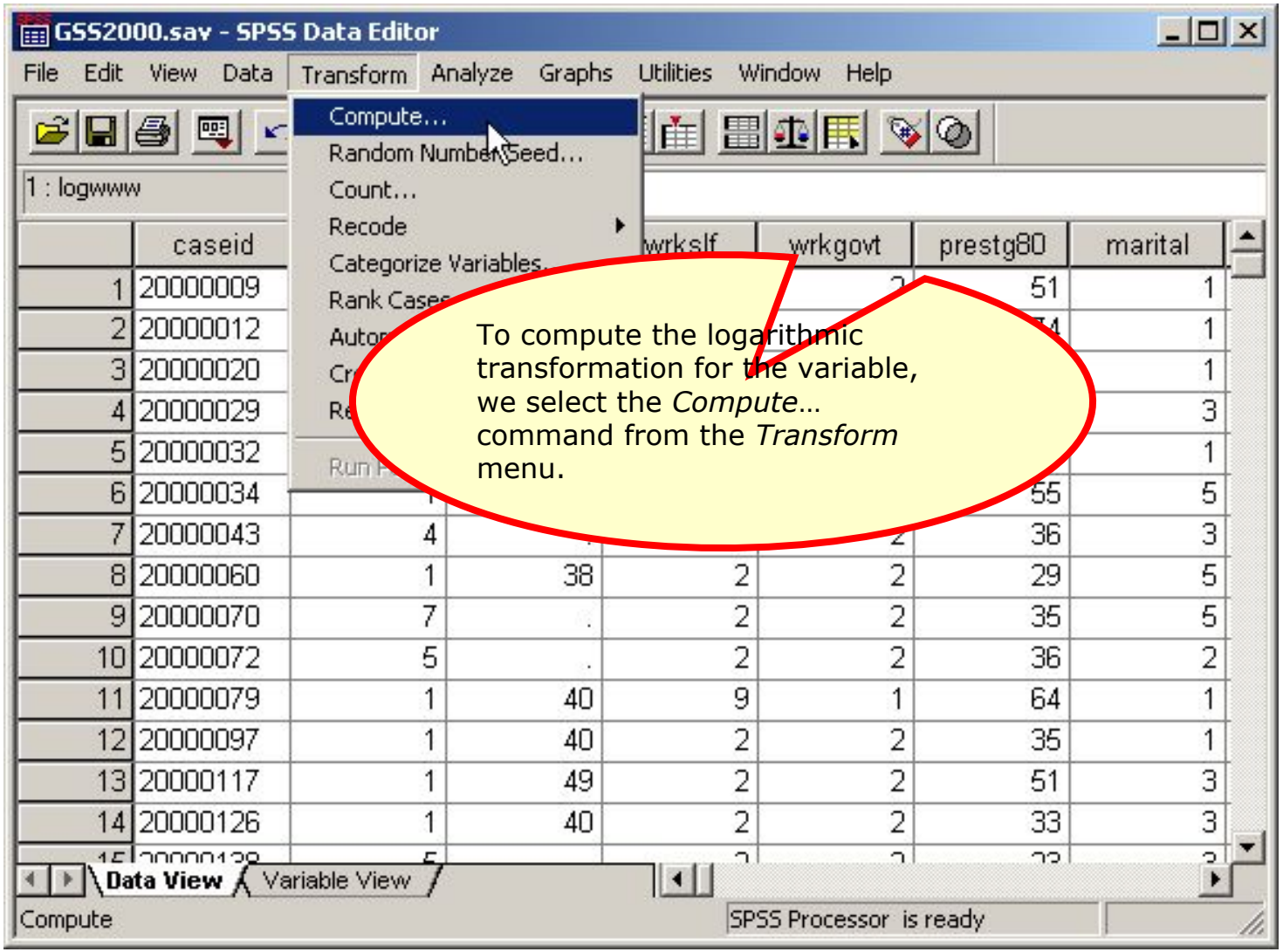
Problem 2

In the dataset GSS2000.sav, is the following statement true, false, or an incorrect application of a statistic?

Based on a diagnostic hypothesis test for homogeneity of variance, the variance in "highest academic degree" is not homogeneous for the categories of "marital status." However, the variance in the logarithmic transformation of "highest academic degree" is homogeneous for the categories of "marital status."

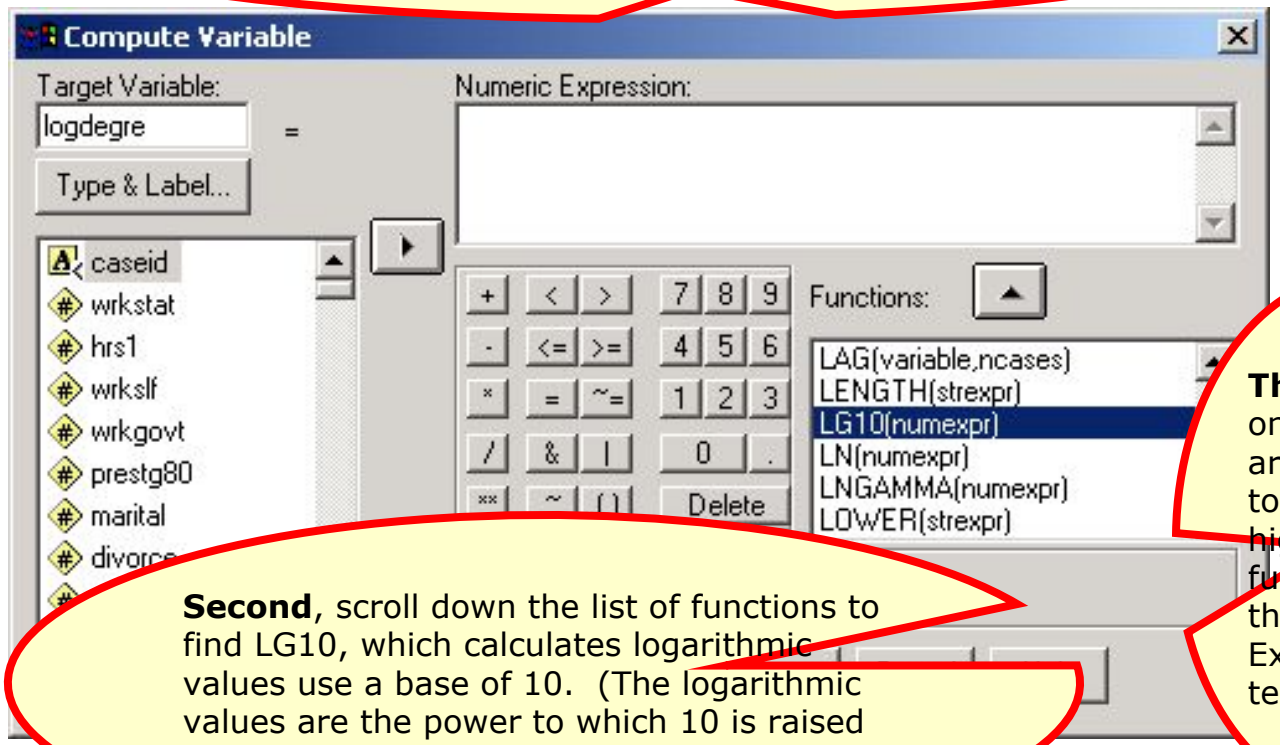
1. True
2. True with caution
3. False
4. Incorrect application of a statistic

Computing the logarithmic transformation



Specifying the variable name and function

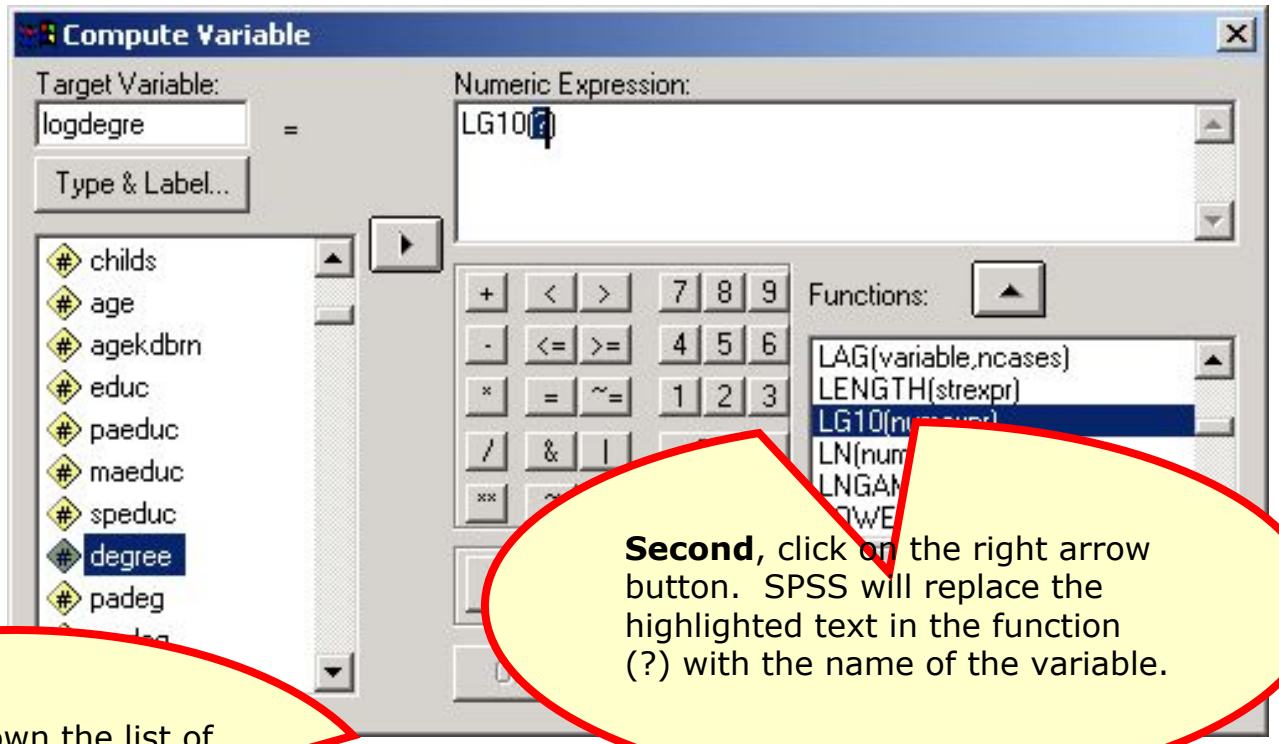
First, in the target variable text box, type the name for the log transformation variable "logdegre".



Second, scroll down the list of functions to find LG10, which calculates logarithmic values use a base of 10. (The logarithmic values are the power to which 10 is raised to produce the original number.)

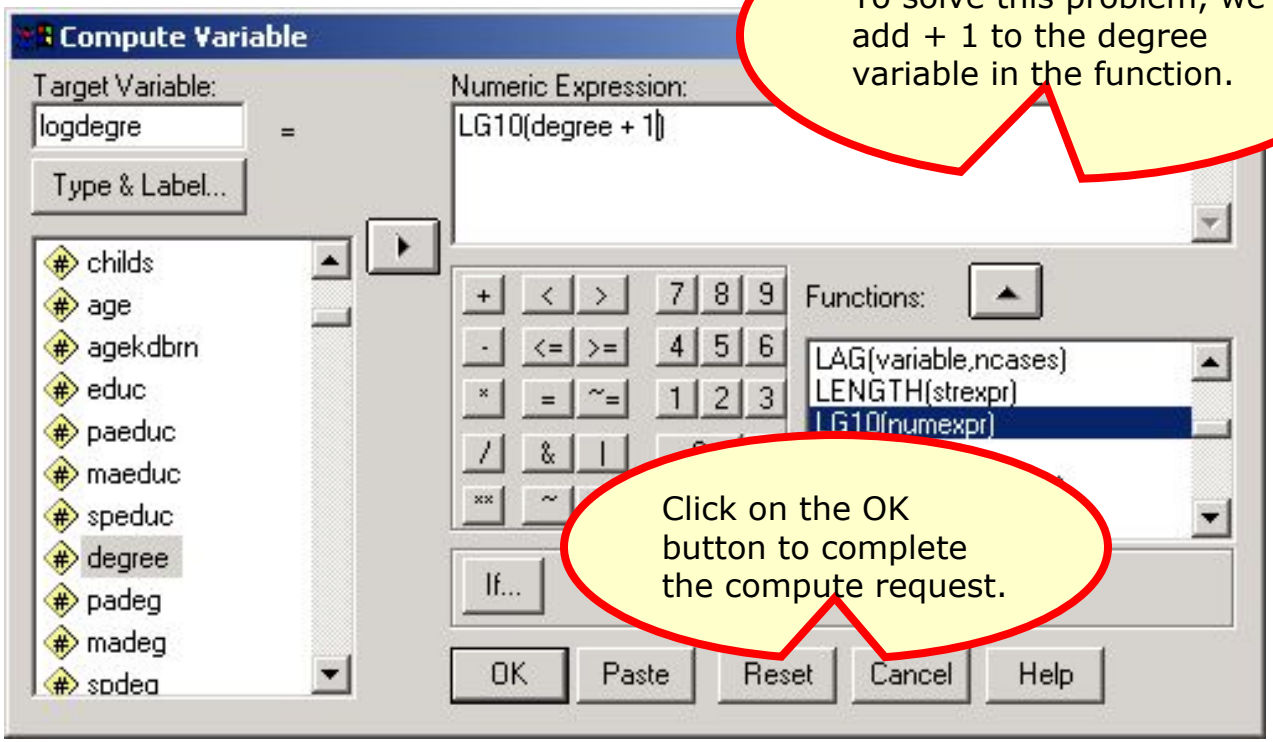
Third, click on the up arrow button to move the highlighted function to the Numeric Expression text box.

Adding the variable name to the function



Preventing illegal logarithmic values

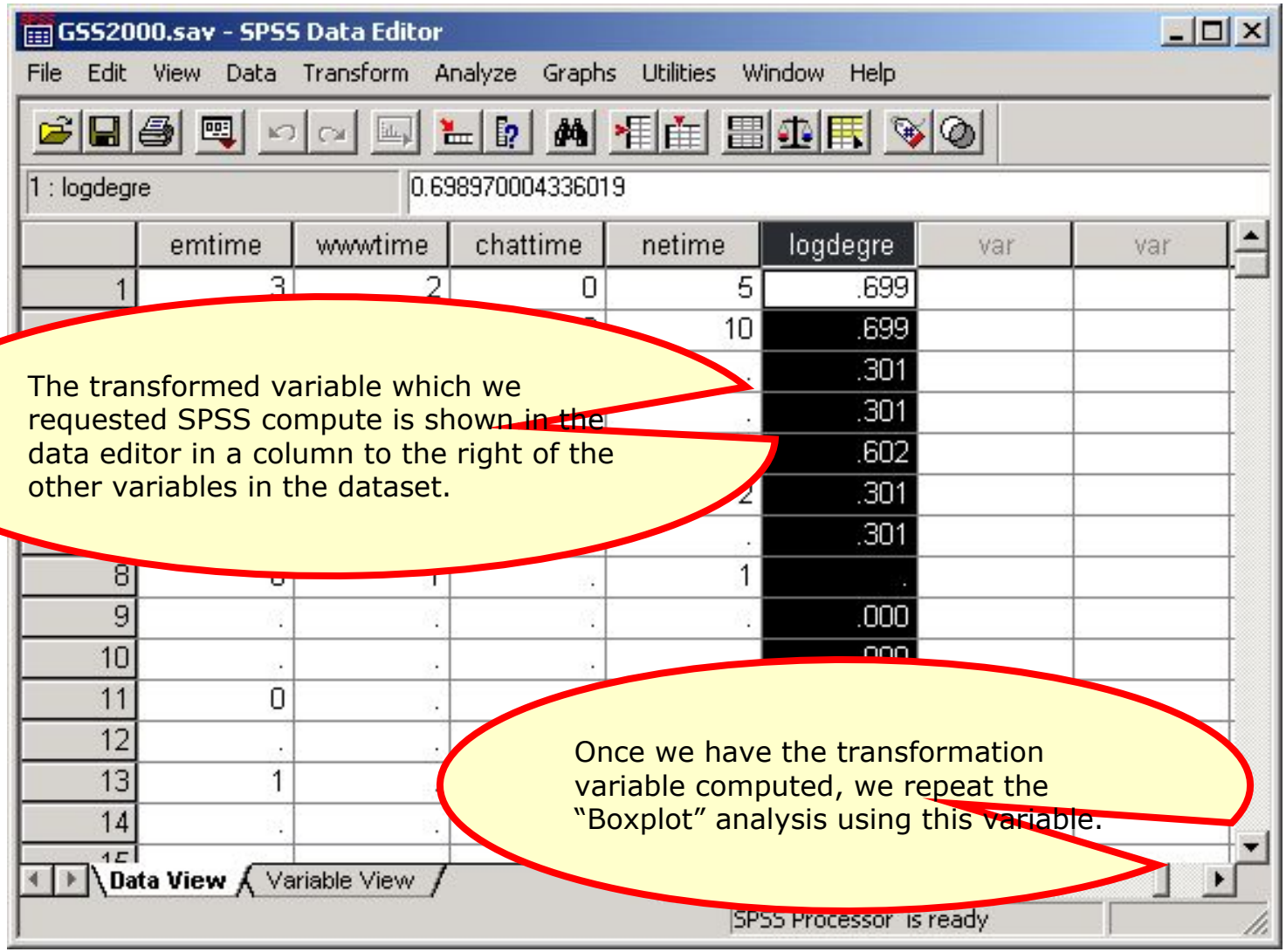
The log of zero is not defined mathematically. If we have zeros for the data values of some cases as we do for this variable, we add a constant to all cases so that no case will have a value of zero.



To solve this problem, we add + 1 to the degree variable in the function.

Click on the OK button to complete the compute request.

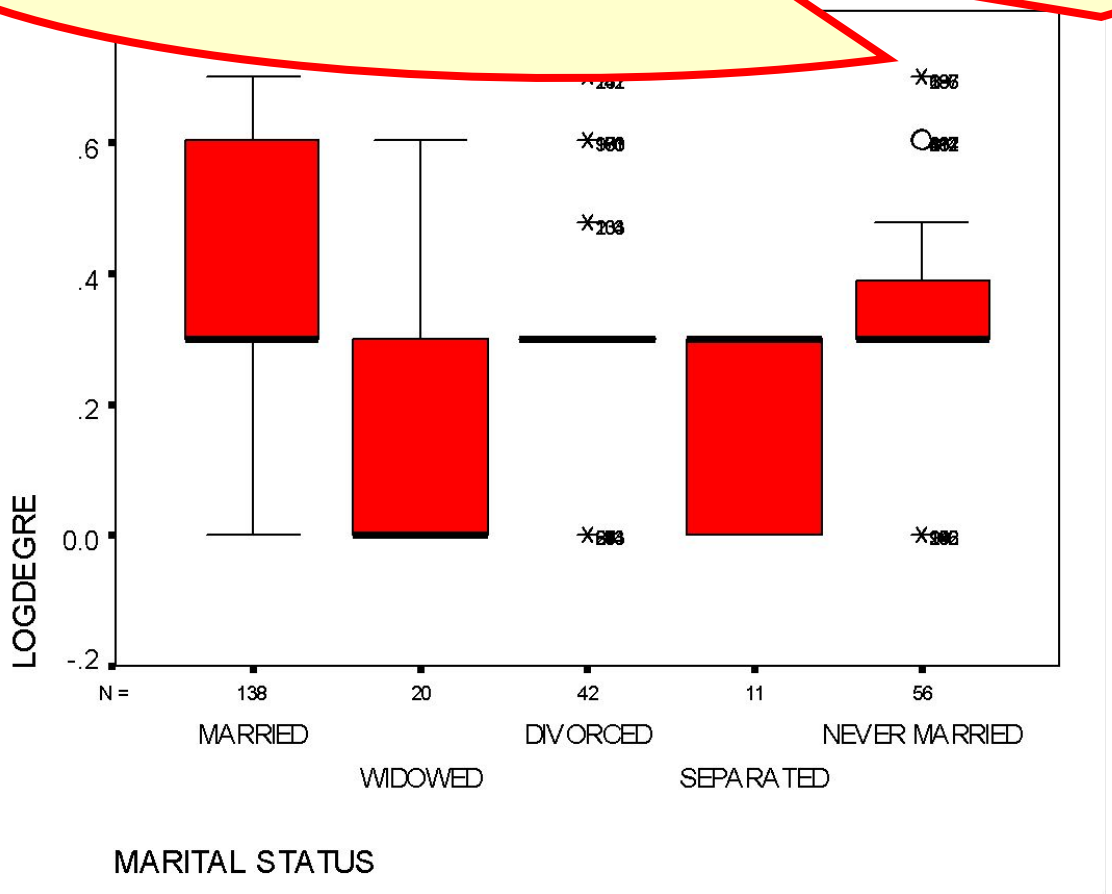
The transformed variable



The boxplot

In this boxplot, the spread is the same for 3 of the 5 groups, which is an improvement over the original boxplot.

However, it is difficult to judge whether or not the problem is solved based solely on the graphic.



The homogeneity of variance test

Test of Homogeneity of Variances

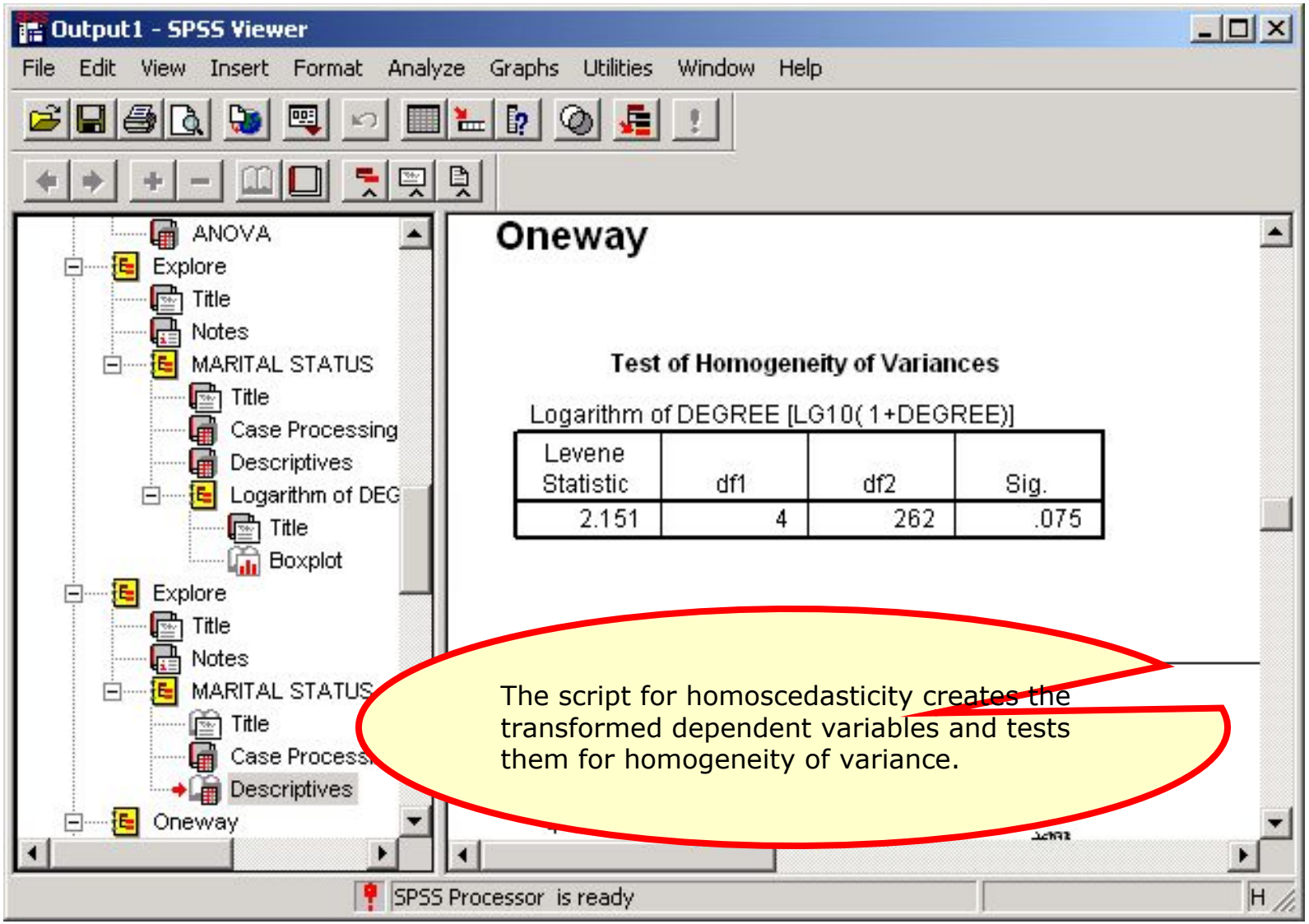
LOGDEGBE			
Levene Statistic	df1	df2	Sig.
.547	4	585	.702

The null hypothesis for the test of homogeneity of variance states that the variance of the transformed dependent variable is equal across groups defined by the independent variable, i.e., the variance is homogeneous.

Since the probability associated with the Levene Statistic (0.075) is greater than the level of significance, we fail to reject the null hypothesis and conclude that the variance is homogeneous.

The answer to the question is **true with caution**.

Homogeneity of variance test from the script



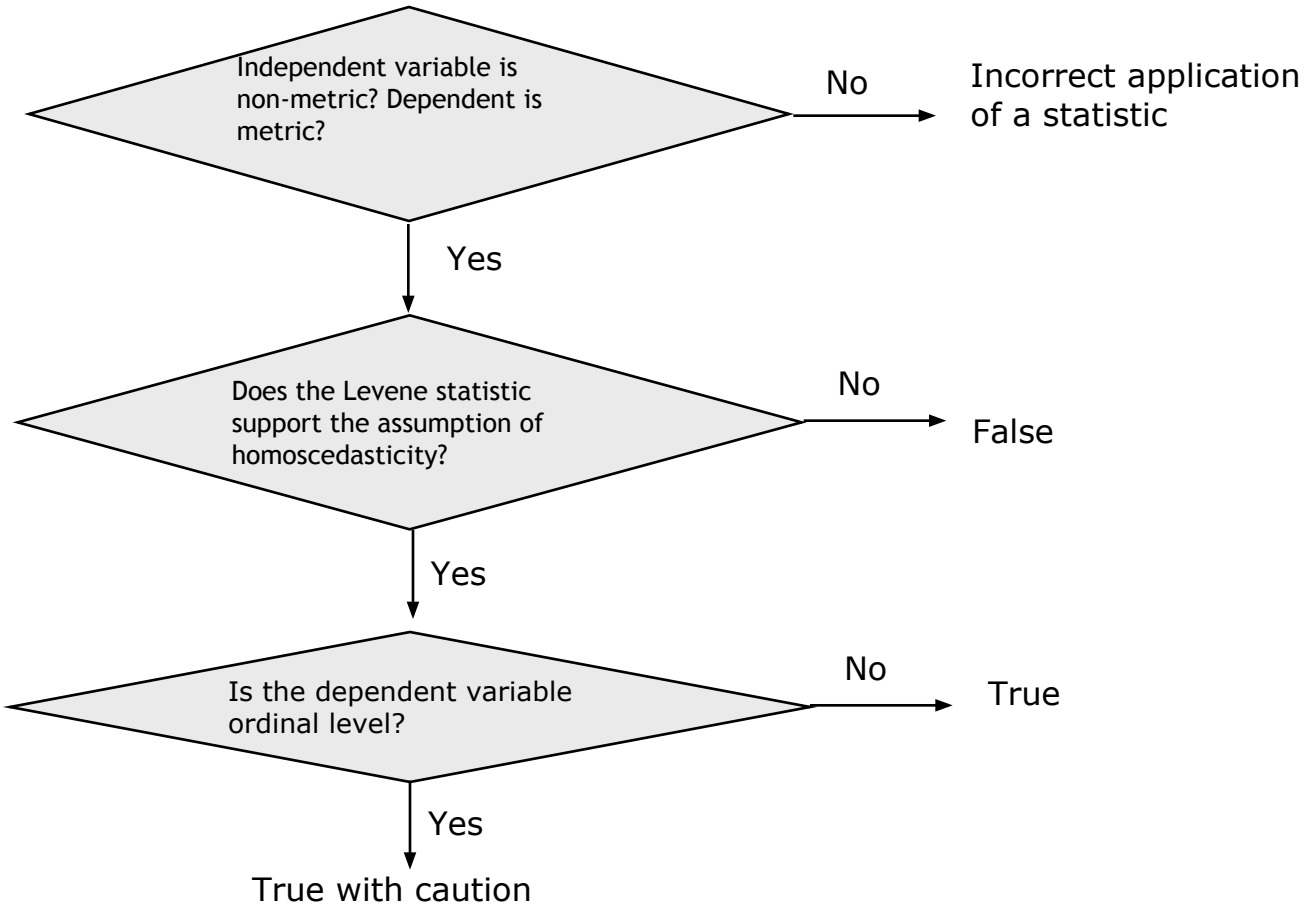
Other problems on homoscedasticity assumption

- A problem may ask about the assumption of homoscedasticity for a nominal level dependent variable. The answer will be "An inappropriate application of a statistic" since variance is not computed for a nominal variable. Similarly, an ANOVA cannot be calculated if the independent variable is interval level and the answer will be "An inappropriate application of a statistic."

- A problem may ask about the assumption of homoscedasticity for an ordinal level dependent variable. If the variable or transformed variable satisfies the assumption of homogeneity of variance, the correct answer to the question is "True with caution" since we may be required to defend treating ordinal variables as metric.

Steps in answering questions about the assumption of homoscedasticity – question 1

The following is a guide to the decision process for answering problems about the homoscedasticity of a variable:



Steps in answering questions about the assumption of homoscedasticity – question 2

The following is a guide to the decision process for answering problems about the homoscedasticity of a transformation: