The Ideal Qualitative Study with a Quantitative Component
6 Week Webinar Class

CLASS 2:
QUANTITATIVE 2

Welcome to week two of our 6 week webinar class *The Ideal Qualitative Study with a Quantitative Component*. The focus of this week is completing the quantitative analysis section of our study titled *Quantitative Analysis 2*:

Week 1: Intro. & Quantitative Analysis Part 1

Week 2: Quantitative Analysis Part 2

Week 3: Essentials of Qualitative Analysis

Week 4: Essentials of Mixed Methods Research

Week 5: Statistical Analysis with Small Samples

Week 6: Write-up and Report

Within our model presenting *The 7 Steps of Data Analysis* (Quantitative Analysis), our focus this week will be Steps 4, 5, and 6, which reflect the statistical analysis portion of this model:

Step 1: Study Map

Step 2: Data Entry

Step 3: Check Data Integrity

Step 4: Univariate Analysis

Step 5: Bivariate Analysis

Step 6: Multivariate Analysis

Step 7: Write-up and Report

Step 4: Univariate Analysis

Univariate analysis explores each variable in a dataset separately. This is reflected in the term univariate analysis as the prefix *uni* suggests the number one and the root *var* suggests variable, indicating the analysis of one variable. Univariate analysis is also referred to as **descriptive analysis**, as these statistics essentially *describe* variables. It may seem logical to conduct an initial univariate analysis where individual variables are described, prior to inferential analysis where variables are related to one another.

Univariate analysis commonly involves reporting measures *of* ***central tendency*** and ***dispersion***. Measures of central tendency involve describing how the values within a distribution of scores *tend* to be clustered *centrally*. The most common measure of central tendency is the **mean** score, which is simply the arithmetic average of a set of scores (where the values of a number of items are summed, then that summed number is divided by the number of items summed). If a distribution of scores in non-normal, the **median** (the middlemost numerical value separating the higher half of a distribution of scores from the lower half) or **mode** (the value that appears most often in distribution of scores) might be a more representative measure of central tendency. If a distribution of scores is perfectly normal, the mean, median, and mode will intersect.

Another set of values presented in univariate analysis, involve measures of ***dispersion***, which describe how scores tend to be dispersed in a distribution. Common measures of dispersion include the **range** (the highest – lowest value), which reflects the distance between the lowest and highest value, as well as the **minimum and maximum** scores (the lowest and highest values in a distribution of scores). A more sophisticated measure of dispersion is the **standard deviation**. The standard deviation portrays the dispersion of values around the mean. Specifically, the standard deviation reflects the average variability (or spread) of the scores from the mean value.

Please see the **Figure 1**, which presents the *Univariate Test Key*, which is an abbreviated key that describes how univariate analysis is typically conducted in a data analysis study. In the first column (labeled **Variable**) we see there are two types of variables that need to be presented via univariate analysis, **Categorical** (nominal and ordinal level variables) and **Continuous** (ratio and interval level variables). In the center column (labeled **Test to Use**) we see the **Frequencies** procedure in SPSS will be used for both types of variables. However, in the final column (**Values to Report**) we see that we will compute and report different statistical values for the categorical

(number and percentage by category) and continuous (mean, standard deviation, and minimum/maximum score).

Figure 1 Univariate Test Key

Variable	Test to Use (Values to Report)
Categorical	Frequencies (Number & Percentage)
Continuous	Frequencies (Mean, Std. Deviation, Min/Max)

Plugging in the Study Variables. Please refer to **Figure 2**, which presents the *Univariate Test Key* with the study variables plugged in. Once we have identified the variables used in the sample study as categorical (*Doctor Prescription Status, Class Assignment*) and continuous (*Age*, *Class Enjoyment*, *Class Attendance*), it is quite easy to plug them into the *Univariate Test Key* toward identifying the statistical values to compute and report using the **frequencies procedure** in SPSS.

Figure 2 Univariate Test Key (Study Variables Plugged In)

Variable	Test to Use (Values to Report)
Categorical	Frequencies (Number & Percentage)
Doctor Rx Status	Frequencies (Number & Percentage)
Class Assignment	Frequencies (Number & Percentage)
Continuous	Frequencies (Mean, Std. Deviation, Min/Max)
Age	Frequencies (Mean, Std. Deviation, Min/Max)
Class Enjoyment	Frequencies (Mean, Std. Deviation, Min/Max)
Class Attendance	Frequencies (Mean, Std. Deviation, Min/Max)

Categorical Variables

Computing univariate statistics for categorical variables is rather simple. In fact, most statistics textbooks often only describe how to report univariate statistics for continuous variables. However, it is useful to keep in mind that when describing a categorical variable, one should present the number and percentage of study participants that fall into each response category. For example, one should present the number and percentage of study participants that fall into the response categories for the covariate variable *Doctor Prescription Status* (Doctor Rx - No, Doctor Rx - Recommended, and Doctor Rx - Yes).

These values are produced in SPSS by going to:

1) **Analyze→ Descriptive Statistics→ Frequencies**.
2) In the dialogue box that opens (displayed in **Figure 4.25**), move the categorical study variables (*Doctor Prescription Status, Class Assignment*) from the left hand column into the right hand column.
3) Click **OK** or **Paste**.

SPSS Output 1 presents the statistical output for the frequencies procedure. The first box labeled **Statistics** reports the number of cases with valid and missing values, which indicates 100 valid cases and 0 missing for each variable, which means both variables have full data and no missing data points. The next box (labeled **Frequency Table**) presents the frequency counts reflecting how many study participants are in each category within each variable. For example, regarding the variable *Doctor Prescription Status*, among the 100 study participants, 40 (40%) are at the Doctor Rx - No level, 30 (30%) are at the Doctor Rx - Recommended level, and 30 (30%) are at the Doctor Rx - Yes level. Regarding the variable *Class Assignment* of the 100 study participants, 50 (50%) are assigned to Morning classes and 50 (50%) are assigned to evening classes.

SPSS Output 1

Frequencies

Statistics

		Doctor Prescription Status	Class Assignment
N	Valid	100	100.0
	Missing	0	0

Frequency Table

Doctor Prescription Status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Doctor Rx - No	40	40.0	40.0	40.0
	Doctor Rx - Recommended	30	30.0	30.0	70.0
	Doctor Rx - Yes	30	30.0	30.0	100.0
	Total	100	100.0	100.0	

SPSS Output 1 (continued)

Class Assignment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Morning	50	50.0	50.0	47.0
	Evening	50	50.0	50.0	100.0
	Total	100	100.0	100.0	

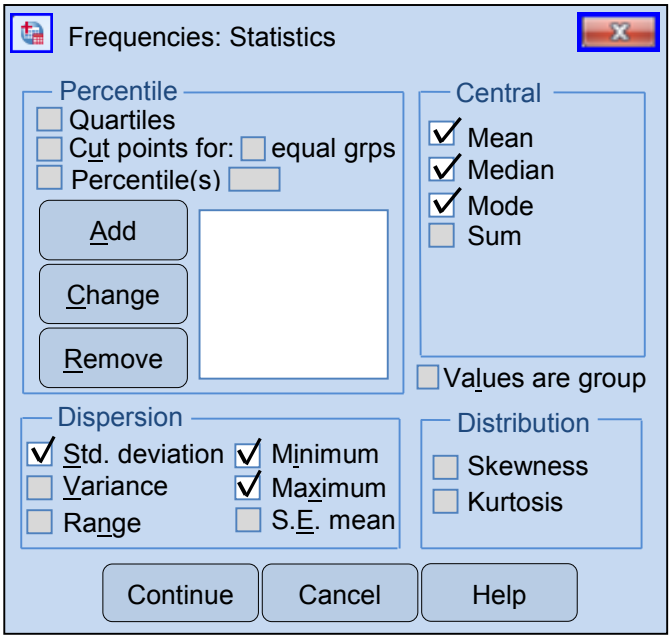
Continuous Variables

As stated earlier, computing univariate statistics for continuous variables largely involves measures of central tendency and dispersion. Recall the *Univariate Test Key* noted the most commonly cited measure of central tendency, which is the *mean* value, as well dispersion, which are the *standard deviation* and *minimum/maximum* values. To produce these univariate statistics for our continuous study variables, in SPSS go to:

1) **Analyze→ Descriptive Statistics→ Frequencies**.
2) In the dialogue box that opens (displayed in **Figure 4.25**) use the center arrow to move the continuous variables (*Age*, *Class Enjoyment*, *Class Attendance*) from the left column into the column on the right.
3) Click **Statistics**.
4) In the dialogue box that opens, click **Mean**, **Median**, **Mode**, (measures of central tendency), as well as **Std. deviation**, **Minimum**, and **Maximum** (measures of dispersion) as presented in **Figure 3**.
5) Click **Continue**.
6) Click **OK** or **Paste**.

SPSS Output 2 presents the statistical output for this frequency procedure. The first box (labeled **Statistics**) in the output reports the number of cases with valid and missing values. Again, this box indicates that for both variables there are complete data for all 100 study participants. That same box presents the values for *Age* reflecting the mean (79.29), median (81.50), Mode (84.00), standard deviation (8.75), and minimum (61.00) /maximum (94.00) values. Additionally, the box presents values for *Class Enjoyment* of mean (3.00), median (3.00), Mode (2.00), standard deviation (1.40), and minimum (1.00) /maximum (5.00) values. Furthermore, in the same box we see the values for *Class Attendance* reflecting the mean (3.69), median (3.80), Mode (3.00), standard deviation (1.81), and minimum (0.00)/maximum (7.00) values.

Figure 3 Activating measures of central tendency and dispersion



SPSS Output 2

Frequencies

Statistics

		Age	Class Enjoyment	Class Attendance
N	Valid	100	100	100
	Missing	0	0	0
Mean		79.2900	2.9950	3.6870
Median		81.5000	3.0000	3.8000
Mode		84.00	2.00	3.00
Std. Deviation		8.75168	1.40089	1.80821
Minimum		61.0	1.00	0.00
Maximum		94.0	5.00	7.00

Step 5: Bivariate Analysis

After employing Univariate Analysis to examine the variables by themselves, we may now want to look at how the variables relate to one another using Bivariate Analysis. Bivariate Analysis involves looking at how two variables at a time relate to one another. Indeed, if we break down the term bivariate we see the roots of the term indicate looking at two variables where the term *bi* indicates two and *var* indicates variable. Recall, we want to include all study variables related to the dependent variable at a statistically significant level ($p<.05$) in the multivariate regression model to statistically control for the influence of those variables in the final stage of analysis. Therefore, a primary goal of bivariate analysis is not just to test the association between two variables, but also identify which study variables are related to the dependent variable, so they may be included as predictors in the final multivariate regression model conducted in **Step 6: Multivariate Analysis**.

For example, in sample study one, a goal of bivariate analysis would be to identify which study predictor variables (*Age, Doctor Prescription Status, Class Enjoyment, Class Assignment*) are significantly related to the dependent variable *Class Attendance* on a one on one level (i.e., bivariate level). Each study predictor variable related to the dependent variable at a statistically significant level, will be included in the multiple linear regression model examining *Class Attendance*. Any predictor variables not significantly related to the dependent variable will be excluded from the multivariate analysis.

Figure 4 presents the *Bivariate Test Key*, which describes the appropriate statistical procedure to test the association between two variables, will depend on the structure of each variable. **Figure 5** presents the *Bivariate Test Key* with the study variables plugged in. The first line instructs if Variable 1 is categorical with a two category response, such as *Class Assignment* (Response: Morning, Evening) and Variable 2 is continuous, such as *Class Attendance* (Response: 1-7), then an **Independent Samples T-Test** would be the appropriate statistical procedure to test this association. The second line within the figure describes that if Variable 1 is categorical with three or more categories, such as *Doctor Prescription Status* (Response: Doctor Rx - No, Doctor Rx - Recommended, or Doctor Rx - Yes) and Variable 2 is continuous, such as *Class Attendance* (Response: 1-7), then a **One-Way ANOVA** would be the appropriate statistical test to examine this association. The third line instructs if Variable 1 is

continuous, such as *Age* (Response: NA) or *Class Enjoyment* (Response: 1-5) and Variable 2 is continuous, such as *Class Attendance* (Response: 1-7), then a **Correlation** would be the appropriate statistical test to examine this association. The chi square analysis listed on the fourth line is not required here, but will be reviewed in detail later.

Figure 4 Bivariate Test Key

Variable 1	Variable 2	Test To Use
Categorical (2 categories)	Continuous	Independent Samples T-Test
Categorical (≥3 categories)	Continuous	One-Way ANOVA
Continuous	Continuous	Correlation
Categorical	Categorical	Chi-Square

Figure 5 Bivariate Test Key (Study Variables Plugged In)

Variable 1	Variable 2	Test To Use
Categorical (2 categories)	Continuous	Independent Samples T-Test
Class Assignment (Morn, Eve)	Class Attend (1-7)	Independent Samples T-Test
Categorical (≥3 categories)	Continuous	One-Way ANOVA
Doctor Rx (N/R/Y)	Class Attend (1-7)	One-Way ANOVA
Continuous	Continuous	Correlation
Age (NA)	Class Attend (1-7)	Correlation
Class Enjoyment (1-5)	Class Attend (1-7)	Correlation
Categorical	Categorical	Chi-Square

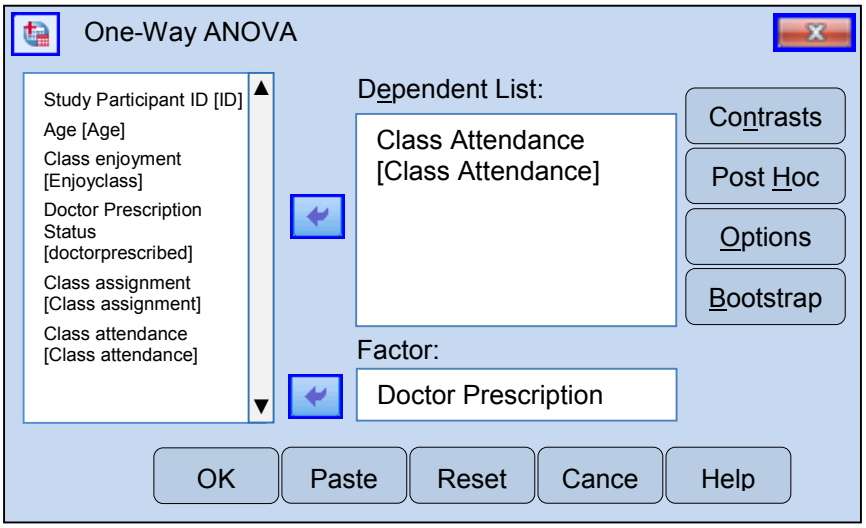
One-Way ANOVA

The Bivariate Test Key indicates a **One-Way ANOVA (Analysis of Variance)** is the appropriate statistical procedure examining the relationship between a categorical variable with three or more response categories with a continuous variable, such as the categorical covariate variable *Doctor Prescription Status* (Response Categories: Doctor Rx - No, Doctor Rx - Recommended, Doctor Rx - Yes) with the continuous dependent variable *Class Attendance* (Response Range: 1-7). This procedure will indicate if there is a statistically significant difference in mean *Class Attendance* scores among study participants with a Doctor Rx - No, Doctor Rx - Recommended, or Doctor Rx - Yes.

The **One-Way ANOVA** procedure has certain assumptions, including **independence of observations** (there is no relationship between the study participants in each group or between the groups themselves, such as there are different participants in each group with no participant being in more than one group), **no significant outliers** (the distributions of scores in the continuous variable should not contain significant outliers scores), **normality** (the scores within the continuous variable are normally distributed), and **homoscedasticity** (homogeneity of variances).

It is important to realize that the number of study participants within each response category within the categorical variable included in the ANOVA is also an important consideration as this will impact statistical power (Kraemer & Thiemann, 1987). In order to detect a medium-to-large effect size, a minimum of *30 participants per cell* has been suggested (Cohen, 1988). As the number of study participants within each response category drops, the power to detect statistically significant differences between groups also wanes. Regarding the **One-Way ANOVA**, 14 study participants per cell will provide approximately 80% power to detect a large effect size (.50), while seven study participants per cell will provide approximately 50% power to detect a large effect size (.50). If a small number of study participants per cell presents an issue, consider a **Kruskal-Wallis *non-parametric ANOVA***, which is also available in SPSS.

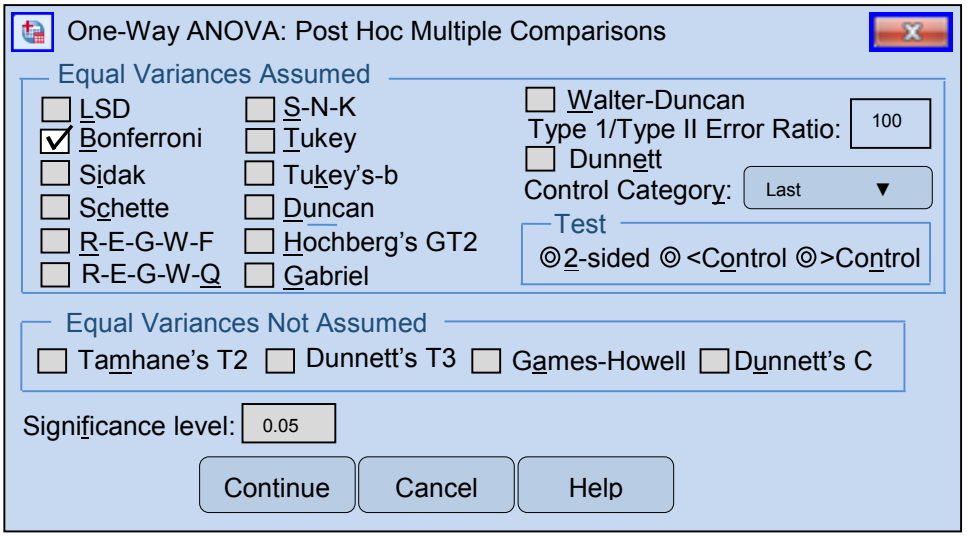
Figure 6 One-Way ANOVA: *Class Attendance* by *Doctor Prescription Status*



To conduct the One-Way ANOVA in SPSS go to:

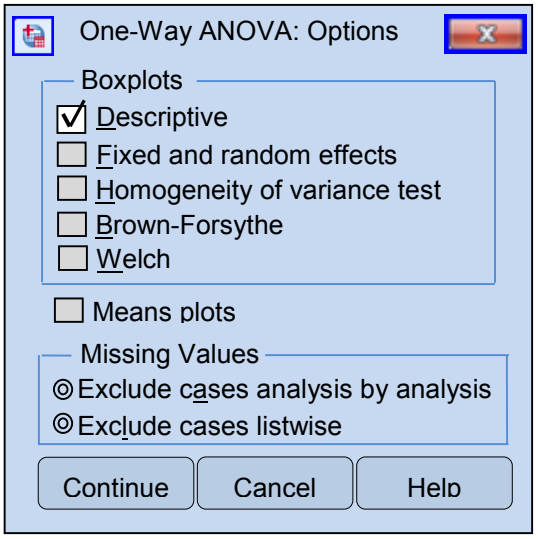
1) **Analyze→ Compare Means→ One-Way ANOVA**.
2) In the dialogue box that opens, enter the continuous variable (the dependent variable *Class Attendance*) from the left hand column into the right labeled **Dependent List**, using the top arrow button.
3) Enter the categorical variable, the covariate predictor variable *Doctor Prescription Status*, into the lower box labeled **Factor** (as presented in **Figure 6**).
4) Click **Post Hoc**.
5) In the dialogue box that opens, click **Bonferroni** (see **Figure 7**).
6) Click **Continue**.

Figure 7 Activating the **Bonferroni post hoc test** in the One-Way ANOVA



7) Click **Options**.
8) In the dialogue box that opens, click **Descriptive**, as presented in **Figure 8**.
9) Click **Continue**.
10) Click **OK** or **Paste**.

Figure 8 Activate the **descriptive** in the One-Way ANOVA



Interpreting the Statistical Output: One-Way ANOVA. **SPSS Output 3** presents the statistical output for the **One-Way ANOVA** procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Class Attendance* and the covariate variable *Doctor Prescription Status*. Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

Note the second box labeled **ANOVA** within the output. In the final column labeled **Sig.**, the level of probability (statistical significance) for the overall test is indicated, which is **.002**. Thus, the overall model for the **One-Way ANOVA** procedure is statistically significant ($p<.05$). Therefore, we will continue to examine the results of the test.

Part B Relationship Dimension 2: Directionality

Next, refer to the top box labeled **Descriptives**, which presents the mean scores and standard deviation values for the dependent variable *Class Attendance* by each response category within the variable *Doctor Prescription Status*. We see that the mean score for *Class Attendance* for study participants with a Doctor Rx - No (M=3.11, SD=1.73) and Doctor Rx - Recommended (M=3.55, SD=1.79), seem to be much lower relative to that of Doctor Rx - Yes (M=4.59, SD=1.61).

Also, refer to the third box labeled **Multiple Comparisons** under the term **Post Hoc Tests**. These values will indicate which mean score is significantly different than another. The first column labeled **(I) Doctor Prescription Status**, presents each category separately. The second column labeled **(J) Doctor Prescription Status**, lists the other response categories. The third column labeled **Mean Difference**, presents the differences between the mean scores relative to the category listed in the first column. The asterisk next to the mean difference value indicates a statistically significant difference, in tandem with the forth column labeled **Sig.**, which presents probability levels.

In the first row of the **Multiple Comparisons** box, we see that mean value *Class Attendance* scores for the study participants with a Doctor Rx - No are significantly lower relative to those with a Doctor Rx - Yes (-1.48333), but Doctor Rx – Recommended is not significantly different. In the next row, we see that mean value *Class Attendance* scores for the study participants with a Doctor prescribed – Recommended are significantly lower than those with a Doctor Rx - Yes (-1.04333). Lastly, the final row describes mean value *Class Attendance* scores for study participants with a Doctor Rx - Yes Doctor Prescription Status, are significantly higher relative to a Doctor Rx - No (1.48333) and Doctor Rx - Recommended (1.04333).

SPSS Output 3

Oneway

(2nd) Directionality:

Higher Doctor Prescription Status = Higher

Descriptives

Class Attendance

	N	Mean	SD	SE	95% CI for the Mean		Min	Max
					Lower	Upper		
Doctor Rx - No	40	3.11	1.73	.27	2.56	3.66	0.0	6.0
Doctor Rx - Rec	30	3.55	1.79	.33	4.22	4.22	1.0	7.0
Doctor Rx - Yes	30	4.59	1.61	.29	5.19	5.19	2.0	7.0
Total	100	3.69	1.81	.18	3.33	4.05	0.0	7.0

ANOVA

(1st) Significance (Model):

Statistically significant (p<.05)

Class Attendance

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	38.52	2	19.26	6.55	.002
Within Groups	285.17	97	2.94		
Total	323.69	99			

SPSS Output 3 (continued)

Post Hoc Tests

(1st) Significance (Mean Scores):
Statistically significant (p<.05)
(Mean score differences)

Multiple Comparisons
Dependent Variable: Class Attendance
Bonferroni

(I) Doctor Prescription Status	(J) Education Level	Mean Difference (I-J)	Std. Error	Sig.	95% CI for the Mean	
					Lower	Upper
Doctor Rx - No	Doctor Rx - Rec	-.44000	.41	.291	-1.26	.38
	Doctor Rx - Yes	-1.48333*	.41	.001	-2.31	-.66
Doctor Rx - Rec	Doctor Rx - No	.44000	.41	.291	-.38	1.26
	Doctor Rx - Yes	-1.04333*	.44	.020	-1.92	-.16
Doctor Rx - Yes	Doctor Rx - No	1.48333*	.41	.001	.66	2.31
	Doctor Rx - Rec	1.04333*	.44	.020	.16	1.92

*The mean difference is significant at the 0.05 level.

Part C

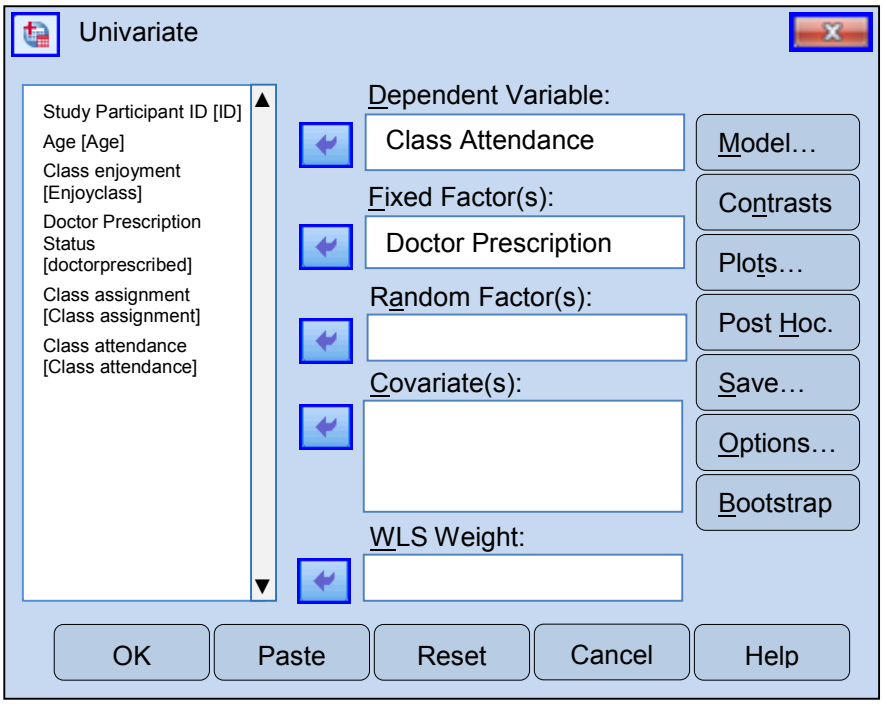
Relationship Dimension 3: Magnitude

The One-Way ANOVA procedure in SPSS does not have an option to estimate the effect size, but the related ANCOVA procedure will produce an effect size estimate known as the partial Eta squared statistic, which can be produced in SPSS through the path:

1) **Analyze→ General Linear Models→ Univariate**.
2) Within the dialogue box that opens, enter the same continuous variable used within the **One-Way ANOVA**, the dependent variable *Class Attendance*, into the top cell under the term **Dependent Variable**.
3) Enter the categorical variable from the **One-Way ANOVA**, the covariate predictor variable *Doctor Prescription Status* into the cell under the term **Fixed Factor(s)**.
4) The dialogue box will look similar to the **Figure 9**.
5) Click **Options**.

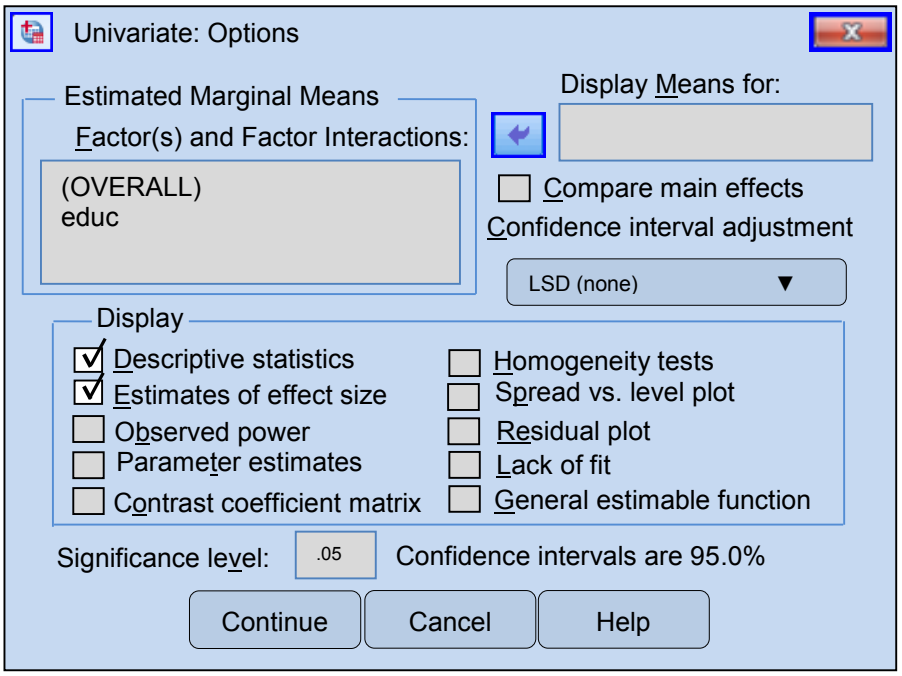
6) Within that dialogue box that opens, click **Estimates of Effect Size** and **Descriptive Statistics**, as presented in **Figure 10**.
7) Click **Continue**.
8) Click **OK** or **Paste**

Figure 9 Producing an effect size using a general linear model



SPSS Output 4 presents the statistical output from the **General Linear Model**. The final column in the box labeled **Tests of Between-Subjects Effects** presents the *Partial Eta Squared (PES)*. In this column, we see that for the variable *doctorprescribed* (*Doctor Prescription Status*) the effect size estimate of the PES is .119. Recall, the effect sizes for the partial Eta squared are small (>0.01), medium (>0.06), and large (>0.14). Thus, the predictor *Doctor Prescription Status* has a medium to large effect size (.119) on the dependent variable *Class Attendance*.

Figure 10 Activating **Estimates of Effect Size** and **Descriptive Statistics**



SPSS Output 4

Tests of Between-Subjects Effects

Dependent Variable: Class Attendance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	38.52	2	19.26	6.55	.002	.119
Intercept	1381.50	1	1381.5	469.9	.000	.829
doctorprescribed	38.52	1	19.26	6.55	.002	.119
Error	285.17	97	2.94			
Total	1683.09	100				
Corrected Total	323.69	99				

(3rd) Magnitude:
Large effect size (PES = .119)

Summary. The One-Way ANOVA indicated a statistically significant relationship between variables as the probability level for this test was less than .05 ($p<.05$). Actually, probability is reported as $p<.01$ as the level of probability was .002 (Sig.=.002). In terms of directionality, the mean scores, as well as the post hoc test indicated the group Doctor Rx - Yes (M=4.59, SD=1.61), was associated with a higher level of *Class Attendance*, relative to Doctor Rx – Recommended (M=3.55, SD=1.79) and Doctor Rx - No (M=3.11, SD=1.73). Lastly, in terms of magnitude, the *partial Eta squared* effect size (.119) indicated that covariate variable *Doctor Prescription Status* evidenced medium to large effect upon the dependent variable *Class Attendance*.

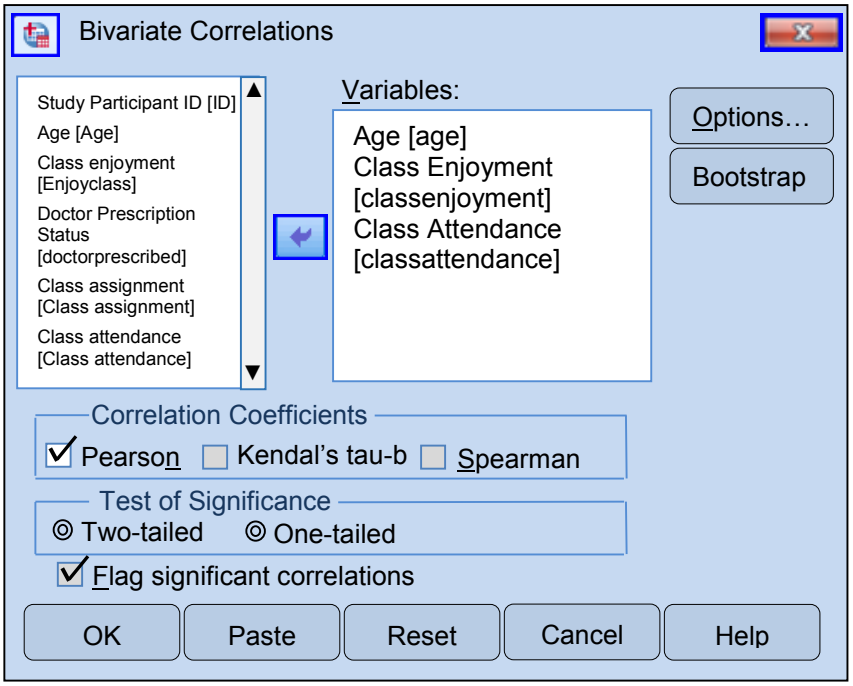
Correlation

The Bivariate Test Key indicates a correlation is the appropriate statistical procedure examining the relationship between a continuous variable with another continuous variable. Therefore, we will use this procedure to examine the relationship between the continuous covariate variable *Age* (Response Range: NA) and *Class Enjoyment* (Response Range: 1-5) with the continuous dependent variable *Class Attendance* (Response Range: 1-7). This procedure will indicate if *Class Attendance* scores either increase or decrease at a statistically significant level as *Ages* and *Class Enjoyment* increase or decrease. Of the several types of tests of correlation, we will perform the **Pearson product-moment correlation** (the most commonly used). If a small sample size is an issue, we would consider a **Spearman's rank *non-parametric correlation***, which is also available in SPSS. The Pearson's correlation has several assumptions, including **linearity** (the relationship between variables must be linear), **normality** (the distribution of scores for both variables must be approximately normal), and **no outliers** (the distributions should not contain significant outliers scores).

To conduct the Pearson's correlation in SPSS go to:

1) **Analyze→ Correlation→ Bivariate**.
2) Then move the two continuous variables, *Age* and *Class Attendance*, from the left into the right hand column as presented in **Figure 11**.
3) Check the box next the term **Pearson**.
4) Click either **OK** or **Paste**.

Figure 11 Pearson's correlation: *Age, Class Enjoyment,* and *Class Attendance*



Interpreting the Statistical Output: Pearson's Correlation. **SPSS Output 5** presents the statistical output for the **Pearson's correlation** procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Class Attendance* and the covariate variables *Age* and *Class Attendance*. Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

Within the single box produced in the statistical output, labeled **Correlations,** the same results are presented in each intersecting cell, so we could refer to either. We will use the cells in the first row that begins with the cell that reads *Class Attendance*. Within this row, note the intersecting columns labeled *Age* and *Class Enjoyment*. Refer to the second line labeled **Sig. (2-tailed)**, which presents the **statistical significance** (probability level) for the procedure. We can see the p-value in the intersecting columns for *Age* and *Class Enjoyment* is .006 and .001, respectively. This indicates both tests are statistically significant (p<.05).

(3rd) Magnitude:
Large effect size (r = -.272)

(2nd) Directionality:
Negative relationship (-)

(3rd) Magnitude:
Large effect size (r = .339)

(2nd) Directionality:
Positive relationship (no -)

(1st) Significance:
Statistically significant

(1st) Significance:
Statistically significant

SPSS Output 5

Correlations

		Class Attendance	Age	Class Enjoyment
Class Attendance	Pearson Correlation	1	-.272**	.339**
	Sig. (2-tailed)		.006	.001
	N	100	100	100
Age	Pearson Correlation	-.272**	1	-676**
	Sig. (2-tailed)	.006		.000
	N	100	100	100
Class Enjoyment	Pearson Correlation	.339**	-676**	1
	Sig. (2-tailed)	.001	.000	
	N	100	100	100

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

Part B Relationship Dimension 2: Directionality

Next, refer to the top row labeled *Pearson Correlation*, which is known as the Pearson *r* correlation coefficient, which will allow us to assess **directionality**. If the correlation coefficient begins with a minus sign, as is the case in the intersecting column labeled *Age*, the relationship with *Class Attendance* is negative. However, if there is not a minus sign, as is the case with the intersecting column labeled *Class Enjoyment*, the relationship with *Class Attendance* is positive. Thus, the findings indicate that as the variable *Age* increases *Class Attendance* decreases. Also, as scores reflecting *Class Enjoyment* increase, *Class Attendance* also increases.

Part C Relationship Dimension 3: Magnitude

The Pearson r correlation coefficient also reflects the **magnitude** or "**effect size**." Specifically, the Pearson r correlation coefficient can range -1 to +1. A value of 0.0 indicates no correlation between variables, but the closer the value is to 1.0 (- or +) the stronger the correlation. A correlation coefficient of .25 is considered a large effect size. We see that the intersecting columns for *Age* and *Class Enjoyment* have a coefficient of -.27 and .34, respectively. Thus, each predictor has evidenced a large effect upon the dependent variable *Class Attendance*.

Summary. Overall, the Pearson's correlation indicated a statistically significant relationship between the dependent variable *Class Attendance* with both predictors *Age* and *Class Enjoyment*. In terms of directionality, the correlation coefficient between *Class Attendance* and *Class Enjoyment* was positive, reflecting as the scores for one variable increases so does the other. The correlation coefficient between *Class Attendance* and *Age* was negative, reflecting as the scores for one variable increases, the other decreases. Lastly, in terms of magnitude, both predictors evidenced a large effect upon the dependent variable *Class Attendance*.

Independent-Samples T-Test

The bivariate test key indicates the **independent-samples t-test** is the appropriate statistical procedure to use when examining a continuous variable with a categorical predictor with two response categories. Therefore, we will use this procedure to examine the relationship between the categorical independent variable *Class Assignment* with the continuous dependent variable *Class Attendance*. This procedure will indicate if there is a statistically significant difference between mean *Class Attendance* scores (Response Range: 1-7), among study participants assigned to morning or evening classes (i.e., *Class Assignment*).

The independent samples t-test procedure has certain assumptions, including **independence of observations** (there is no relationship between the study participants in each group or between the groups themselves, such as there are different participants in each group with no participant being in more than one group), **no significant outliers** (the distributions of scores in the continuous variable should not contain significant outliers scores), **normality** (the scores within the continuous variable are normally distributed), and **homoscedasticity** (homogeneity of variances).

Again, as in the One-Way ANOVA, it is important to consider the number of study participants within each response category within the categorical variable employed in the statistical procedure as this relates to statistical power (Kraemer & Thiemann, 1987). Here, a minimum of *30 participants per cell* has been suggested to detect a medium-to-large effect size (Cohen, 1988). If small numbers of study participants per cell presents an issue, you may want to consider a **Mann-Whitney U Test *non-parametric Independent-Samples test* equivalent**, which is also available in SPSS. To conduct the independent Samples t-test in SPSS go to:

1) **Analyze→ Compare Means→ Independent Samples T-Test**.
2) Move the continuous variable (*Class Attendance*) from the left hand column into the cell on the right labeled **Test Variable(s)**.
3) Move the categorical variable (*Class Assignment*) from the left hand column into the cell on the right labeled **Grouping Variable**.
4) Click **Define Groups**.
5) In the dialogue box that opens enter the value for the first response category, such as a '1' for *Morning* in the **Group 1** cell, then value for the second response category such as '2' for *Evening* in the cell for **Group 2** (see **Figure 12**).
6) Click **Continue**.
7) The original dialogue box will look similar to **Figure 13**.
8) Click **OK** or **Paste**.

Figure 12 Define groups: *Do You Class Assignment*

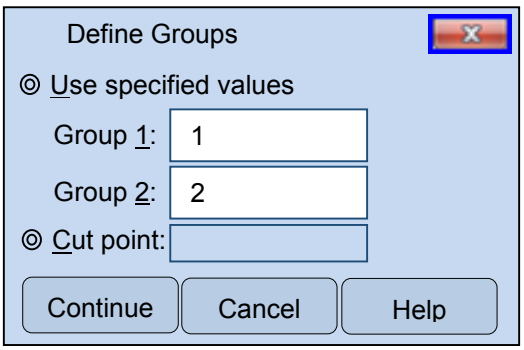
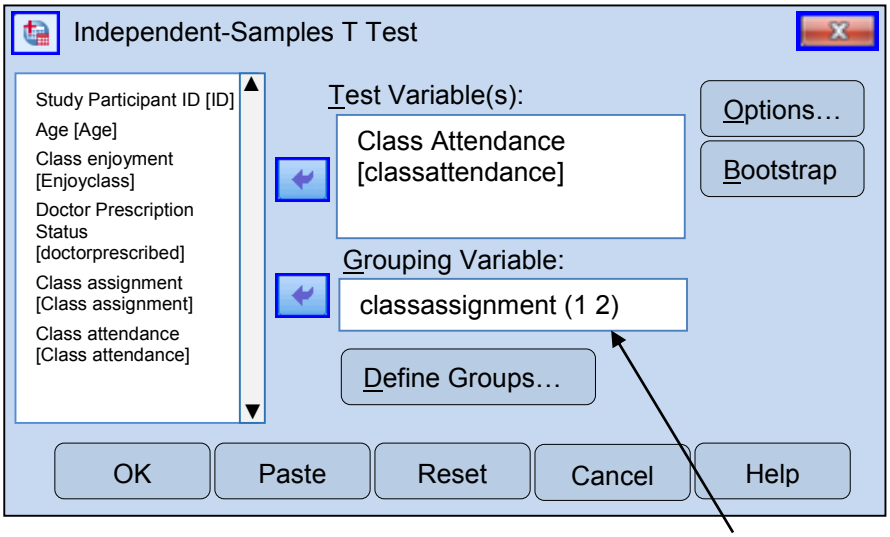


Figure 13 Independent-samples t-test: *Class Attendance* by *Class Assignment*



Assigned to Class in the Morning (=1) or Evening (=2)

Interpreting the Statistical Output: Independent-Samples T-Test. **SPSS Output 6** presents the statistical output for the independent-samples t-test procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Class Attendance* and the independent variable *Class Assignment* Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

Within the statistical output, the second box labeled **Independent-Samples T-Test** presents the level of statistical significance for the procedure. First refer to the column to the left labeled **Levene's Test for Equality of Variances**. The subcolumn beneath labeled **Sig.**, will indicate if the Levene's Test is statistically significant ($p<.05$). If the Leven's test is statistically significant it indicates that the variances in scores reflecting the continuous variable (*Class Attendance*) are significantly unequal among the two categorical response groups (those assigned to morning or evening class). Specifically, if the value in the subcolumn Sig. is .05 ($p<.05$) or below, the values in the second row labeled **Equal Variances not assumed** should be used. In this analysis Sig.=.67, indicating the Levene's test is not statistically significant ($p>.05$) and the top row of

values (**Equal Variances assumed**) should be used. On this line under the column labeled **Sig. (2-tail) is .000** (which reflects the level of statistical significance for the t-test), **indicating that the difference in mean scores is statistically significant (*p*<.05).**

Part B Relationship Dimension 2: Directionality

Top box in the output (labeled **Group Statistics**) presents the mean *Class Attendance* scores for the two categorical groups included in the t-test procedure (study participants that are assigned to *Morning* or *Evening Classes*). Within the column labeled *Class Assignment* the first row labeled *Morning* reflects there are 50 study participants assigned to Morning classes with a mean *Class Attendance* score of 2.89 (*SD*=1.73). The second row labeled *Evening* reflects there are 50 study participants that attended class in the evening with a mean *Class Attendance* score of 4.48 (*SD*=1.52). Thus, the independent-samples t-test indicates that higher ***Class Attendance*** **scores are associated with being assigned to evening classes**.

SPSS Output 6

(2nd) Directionality: Attended Evening = Higher *Class Attendance*

(3rd) Magnitude: Large Effect size (Cohen's *d*= -0.98)

(1st) Significance: Statistically Significant (*p*<.05)

T-Test

Group Statistics

	Class Assignment	N	Mean	Standard Deviation	Std. Error Mean
Class Attend	Morning	50	2.8940	1.73387	.24521
	Evening	50	4.4800	1.52195	.21524

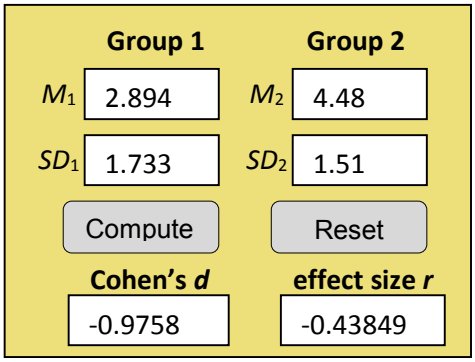
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig	t	df	Sig. (2-tail)	Mean Diff.	95% CI of the Difference	
								Low	Up
Class Attend	Equal Variances assumed	.18	.67	-4.9	98	.000	-1.6	-2.2	-.94
	Equal Variances not assumed			-4.9	96.4	.000	-1.6	-2.2	-.94

Part C Relationship Dimension 3: Magnitude

A common measure of effect size when using an independent-samples t-test is the Cohen's *d*. While the Cohen's *d* is not computed via the independent-samples t-test in SPSS, computing this value is quite straightforward. A simple means of computing the Cohen's *d* is to use one of the many calculators available free online (e.g., http://www.uccs.edu/~lbecker/), which will compute this formula for you. For example, **Figure 14** presents a free online calculator with the mean and standard deviation values from our independent-samples t-test entered for **Group 1** and **Group 2**. Once the **Compute** button has been clicked the Cohen's *d* effect size is produced, which for our analysis is 0.759765 (which matches our manual computation). Recall, the small, medium, and large Cohen's *d* effect sizes are 0.2, 0.5, and 0.8, respectively. Thus, our *d* of 0.7598 reflects an approximately large effect size between the two variables. Specifically, we would say that the independent variable *Class Assignment* has a large effect on the dependent variable *Class Attendance*.

Figure 14 Free online calculator used to compute the Cohen's *d*



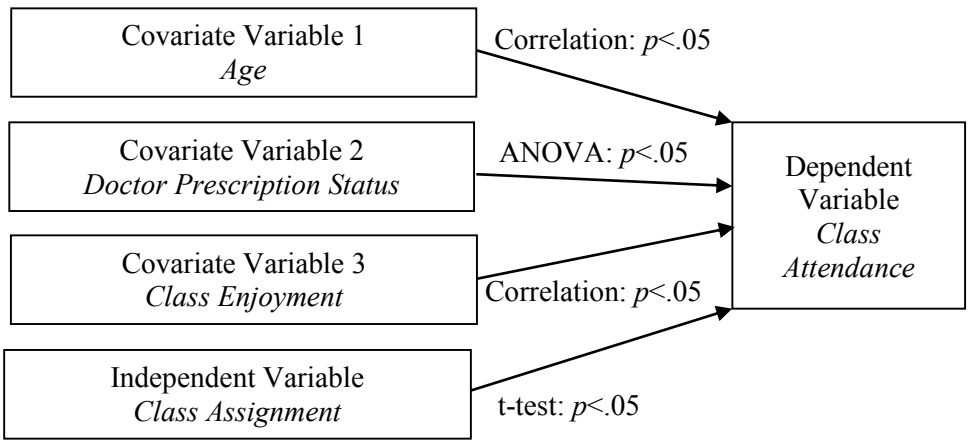
Summary. Overall, the independent-samples t-test indicated a statistically significant relationship between variables (i.e., a difference between mean scores) as the probability level for this test was less than .05 ($p<.05$). Actually, probability is reported as less than .001 ($p<.001$) as the level of probability was .000 (Sig.=.000). In terms of directionality, a higher level of *Class Attendance* was associated assignment to evening class (M=4.48, SD=1.51) relative to a morning class (M=2.89, SD=1.73). Lastly, in terms of magnitude, the Cohen's *d* reflected that the independent variable *Class Assignment* evidenced a large effect (Cohen's *d* effect size=-0.9758) upon the dependent variable *Class Attendance*.

What Did Bivariate Analysis Tell Us?

Within the context of sample study one, our primary goal in bivariate analysis is to identify which predictor variables (covariate/independent variables) have a statistically significant relationship with the dependent variable so these predictors may be included in **Step 6: Multivariate Analysis**. Subsequently, although we examined all three dimensions of the relationships between variables (significance, directionality, and magnitude) in our bivariate testing, in this section we will only review the dimension of statistical significance.

Figure 15 reflects the results of the four bivariate tests used to examine which predictor variables have a statistically significant association with the continuous dependent variable *Class Attendance* (1-7). Notice that we have entered an arrow connecting covariate variable 2: *Doctor Prescription Status* with the dependent variable, which indicates a One-Way ANOVA revealed a statistically significant relationship ($p<.05$). Likewise, there are arrows reflecting statistically significant relationships (correlations) between covariate variable 1: *Age* and covariate variable 3: *Class Enjoyment* with the dependent variable. Lastly, there is an arrow reflecting a statistically significant relationship (via an Independent Samples t-test) between the independent variable: *Class Assignment* and the dependent variable *Class Attendance*.

Figure 15 What Bivariate Analysis Tells Us for in explaining *Class Attendance*



Subsequently, the findings produced through bivariate analysis tells us that all four predictors are significantly related to the dependent variable, and therefore will be included in the next step of multivariate analysis. Please also note, at this step we can only see which predictors are significantly related to the dependent variable, but we cannot see which variable is the strongest predictor of *Class Attendance* (1-7). That dimension will be revealed in multivariate analysis.

STEP 6: MULTIVARIATE ANALYSIS

If we break down the terms within *multivariate analysis* we see the roots of the term suggest examining the relationship between multiple (*multi*) variables (*var*). This is appropriate as we will examine the relationship between *multiple* predictor variables with the dependent variable. Specifically, recall in the previous step we employed bivariate analysis to identify which predictor variables (covariate and independent variables) are related to the dependent variable at a statistically significant level ($p<.05$). In this step we will employ multivariate analysis to identify which predictor variable (*Age, Doctor Prescription Status*, *Class Enjoyment*, *Class Assignment*) is the strongest predictor of the dependent variable *Class Attendance*, which was the overall goal of the research question.

The Multivariate Test Key presented in **Figure 16**, describes that selecting the appropriate multivariate statistical procedure depends upon the structure (i.e., categorical or continuous) of the dependent variable. **Figure 17** presents the **Multivariate Test Key with the study variable plugged in**. Specifically, we see the key with the dependent variable plugged in. The key instructs that the appropriate regression model to use when examining a continuous dependent variable, such as *Class Attendance* (Response Category: 1-7) is a multiple linear regression. **When linear regression involves more than one predictor, the procedure is termed *multiple linear regression*, which is the term we use for this procedure**.

Multiple linear regression has several test assumptions including **linearity**, **normality**, **no undue influence of outlier scores**, **independence of observations, homoscedasticity**, as well as other necessary considerations.

Figure 16 The Multivariate Test Key

Dependent Variable	Regression to Use
Categorical (≥3 categories)	Multinomial Logistic (Not Presented here)
Categorical (2 categories)	Binary Logistic
Continuous	Linear

Figure 17 Multivariate Test Key with the study variable plugged in

Dependent Variable	Regression to Use
Categorical (≥3 categories)	Multinomial Logistic (Not Presented here)
Categorical (2 categories)	Binary Logistic
Continuous	Linear *(>1 Predictor = Multiple Linear)*
Continuous	Class Attendance (1-7)

Dummy-Coding Variables

Dummy-Coding is the key to using a categorical variable with three or more response categories in a regression model. Before the analysis is performed, an important consideration must be made if a categorical predictor variable with three or more response categories is to be included in the regression model. Specifically, if such a categorical variable will be used in the regression model, each response category within the original categorical predictor variable must be **dummy-coded** into a separate **Yes** or **No** dichotomous categorical variable.

For example, in the regression model examining *Class Attendance* the categorical predictor variable *Doctor Prescription Status* will be included, which has three response categories (Doctor Rx - No, Doctor Rx - Recommended, and Doctor Rx - Yes). To use the variable *Doctor Prescription Status* in the regression model, each response category must be made into a separate Yes or No dichotomous categorical variable, which we will describe below. First, consider the original single variable *Doctor Prescription Status* along with the three response categories (note in SPSS data base response categories are called *Values*).

Recall, when we entered these data into the SPSS database, we ascribed the categories the numbers 1, 2, and 3, as presented below:

Response Category (Value) 1: Doctor Rx – No (Coded as "1" in SPSS)
Response Category (Value) 2: Doctor Rx - Recommended (Coded as "2" in SPSS)
Response Category (Value) 3: Doctor Rx - Yes (Coded as "3" in SPSS)

When we dummy-code the variable *Doctor Prescription Status*, we create **three new Yes/No variables** to use in the regression model in place of the original single variable. We will now create these new variables in SPSS using the **Transform** function.

Creating Dummy Coded Variable #1: Doctor Rx – No. First, we will dummy-code to produce Variable 1 ***prescribedno***, to identify study participants that did not receive a prescription from their doctor to take meditation classes. To do this in SPSS, go to:

1) **Transform→ Recode Into Different Variables→.**
2) In the dialogue box that opens, move the variable *Doctor Prescription Status* from the left hand column into the right hand column using the center arrow.
3) Under the box labeled **Name**, enter the name of the first new dummy-coded variable, ***prescribedno***.
4) Click **Change**, which will bring the new variable name into the center column.
5) In the box marked **Label**, enter a description of the new variable (e.g., Doctor Prescription Status: Doctor Rx - No) (**Figure 18** presents the dialogue box).
6) Click the box labeled **Old and New Values**.
7) In the dialogue box that opens, click **Value** in upper left hand column (under **Old Value**) and enter a '1' in the cell (1= ***Doctor Rx – No*** within *Doctor Prescription Status*).
8) Click **Value** in upper right hand column (under **New Value**) and enter a '1' (1=Yes) in the cell, click **Add**.
9) Click **Value** in upper left hand column (under **Old Value**) and enter a '2' in the cell (2=Doctor Rx - Recommended within *Doctor Prescription Status*).
10) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add**.

Figure 18 Dummy-coding the category *Doctor Rx - No*

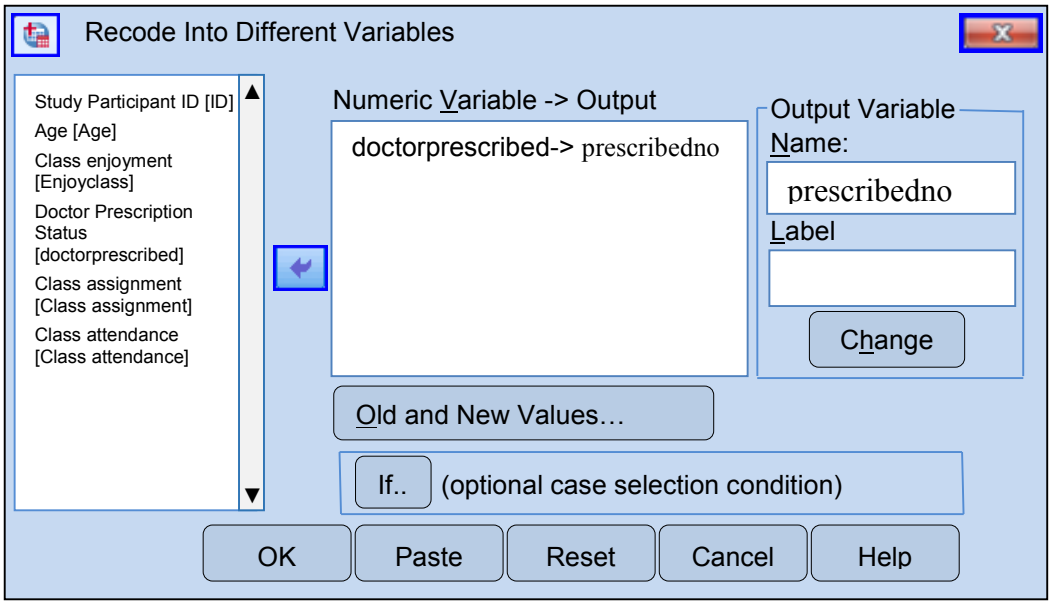
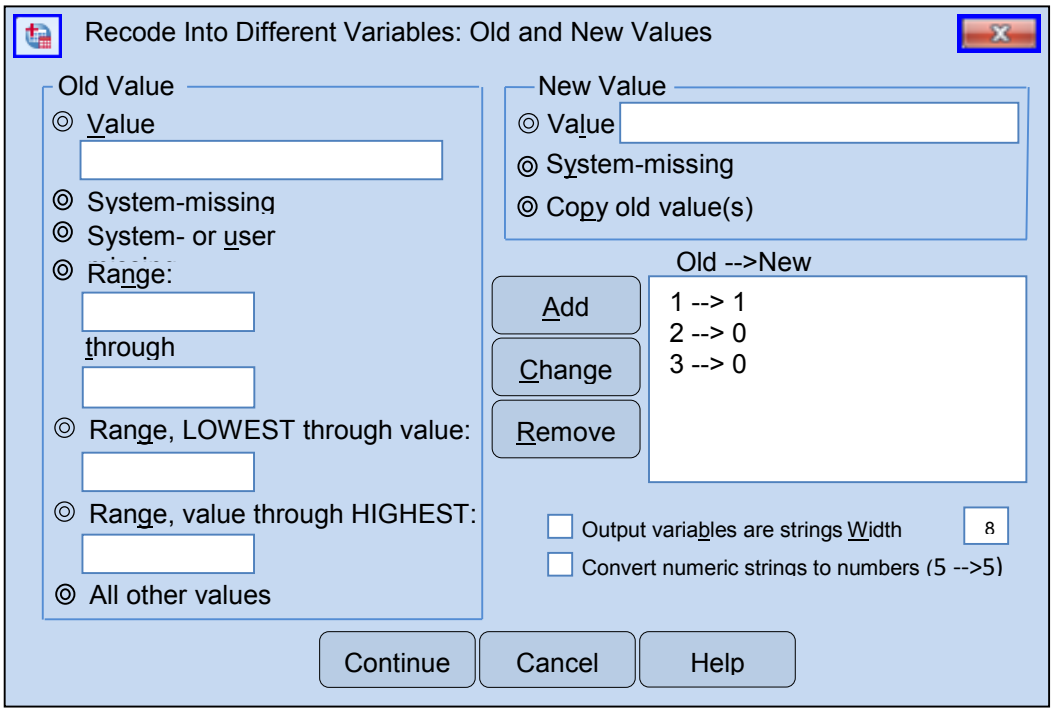


Figure 19 Coding for the dummy-coded category *Doctor Rx - No*



11) Click **Value** in upper left hand column (under **Old Value**) and enter a '3' in the cell (3=Doctor Rx - Yes within *Doctor Prescription Status*).

12) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add** (**Figure 19** presents the dialogue box).

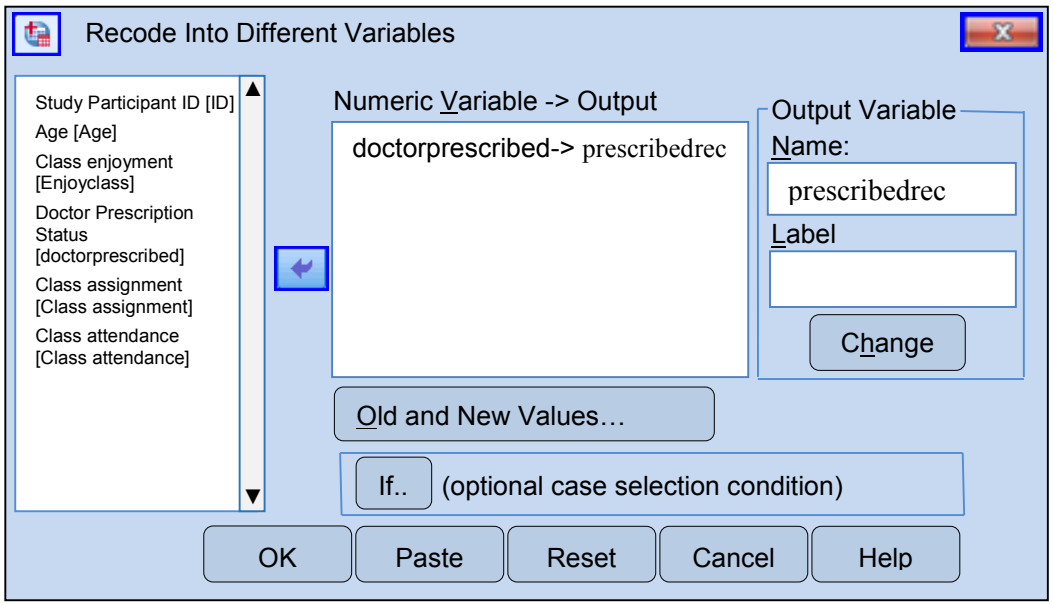
13) Click **Continue**

14) Click **OK** or **Paste**.

Creating Dummy Coded Variable #2: Doctor Rx - Recommended. Next, we will dummy code to produce Variable 2 ***prescribedrec***, which will indicate the study participant that had their doctors recommend they take the meditation classes. To do this in SPSS go to:

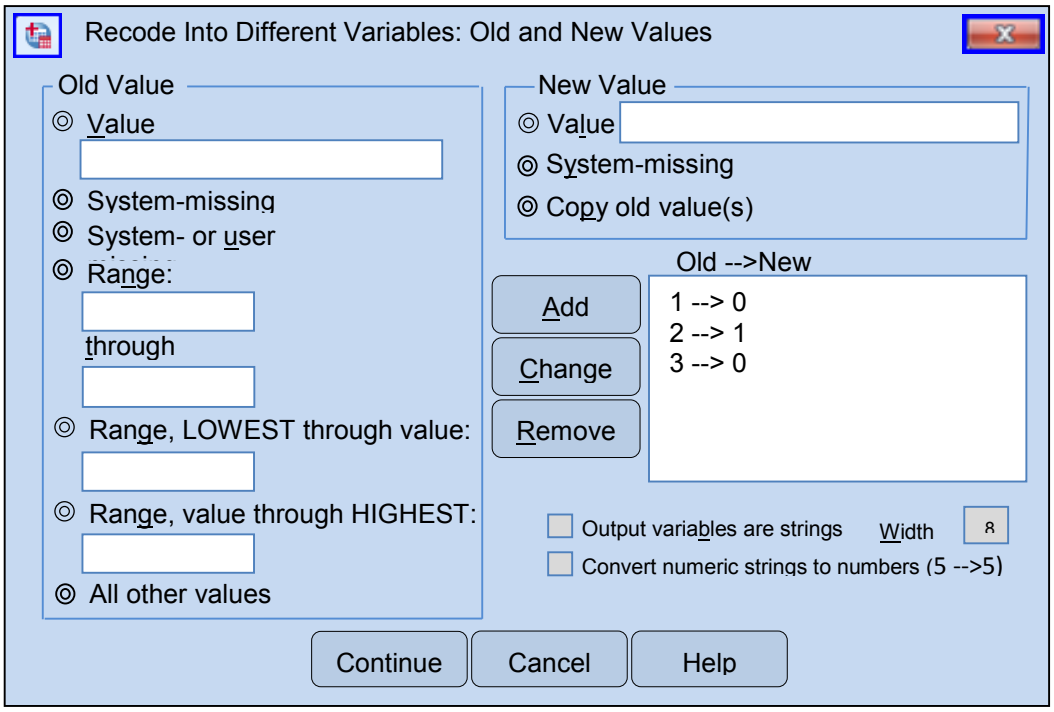
1) **Transform→ Recode Into Different Variables**.
2) In the dialogue box that opens, move the variable *Doctor Prescription Status* from the left hand column into the right hand column using the center arrow.
3) Under the box labeled **Name**, enter the name of the second new dummy-coded variable, ***prescribedrec***.
4) Click **Change**, which will bring the new variable name into the center column.

Figure 20 Dummy-coding the category *Doctor Rx - Recommended*



5) In the box marked **Label**, enter a description of the new variable (e.g., Doctor Prescription Status: Doctor Rx - Recommended) (presented in **Figure 20**).
6) Click the box labeled **Old and New Values**.
7) In the dialogue box that opens, click **Value** in upper left hand column (under **Old Value**) and enter a '1' in the cell (1= ***Doctor Rx – No*** within *Doctor Prescription Status*).
8) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add**.

Figure 21 Coding for the dummy-coded category *Doctor Rx - Recommended*



9) Click **Value** in upper left hand column (under **Old Value**) and enter a '2' in the cell (2=Doctor Rx - Recommended within *Doctor Prescription Status*).
10) Click **Value** in upper right hand column (under **New Value**) and enter a '1' (1=Yes) in the cell, click **Add**.
11) Click **Value** in upper left hand column (under **Old Value**) and enter a '3' in the cell (3=Doctor Rx - Yes within *Doctor Prescription Status*).

12) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add** (presented in **Figure 21)**.

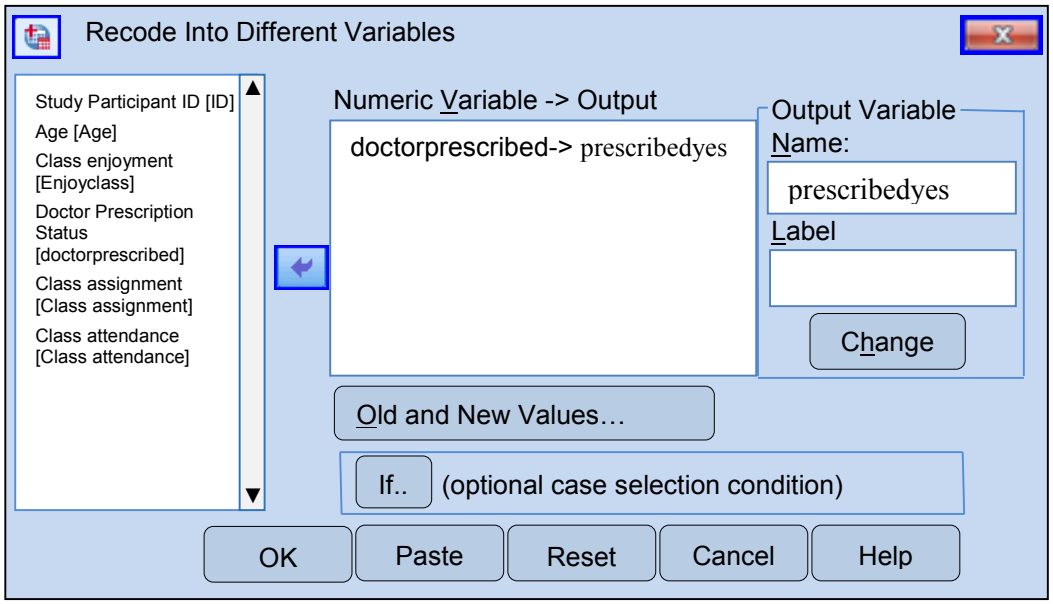
13) Click **Continue**.

14) Click **OK** or **Paste**.

Creating Dummy Coded Variable #3: Doctor Rx - Yes. Next, we will dummy code to produce Variable 3 ***prescribedyes***, which is the variable that will indicate the study participants that had their doctors prescribe meditation class. To do this in SPSS, go to:

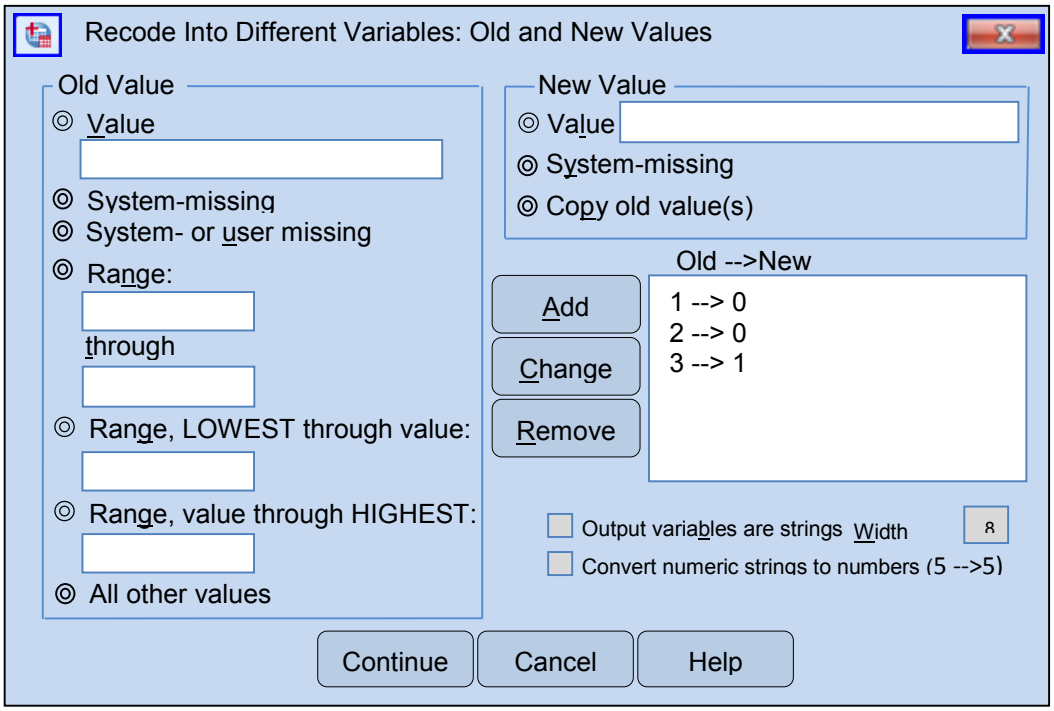
1) **Transform→ Recode Into Different Variables**.
2) In the dialogue box that opens, move the variable *Doctor Prescription Status* from the left hand column to the right hand column using the center arrow.
3) Under the box labeled **Name**, enter the name of the third new dummy-coded variable, ***prescribedyes***.
4) Click **Change**, which will bring the new variable name into the center column.
5) In the box marked **Label**, enter a description of the new variable (e.g., Doctor Prescription Status: Doctor Rx - Yes) presented in **Figure 22**).

Figure 22 Dummy-coding the category *Doctor Rx - Yes*



6) Click the box labeled **Old and New Values**.
7) In the dialogue box that opens, click **Value** in upper left hand column (under **Old Value**) and enter a '1' in the cell (1= Doctor Rx - No within *Doctor Prescription Status*).
8) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add**.
9) Click **Value** in upper left hand column (under **Old Value**) and enter a '2' in the cell (2=Doctor Rx - Recommended within *Doctor Prescription Status*).
10) Click **Value** in upper right hand column (under **New Value**) and enter a '0' (0=No) in the cell, click **Add**.
11) Click **Value** in upper left hand column (under **Old Value**) and enter a '3' in the cell (3=Doctor Rx - Yes within *Doctor Prescription Status*).
12) Click **Value** in upper right hand column (under **New Value**) and enter a '1' (1=Yes) in the cell, click **Add** (presented in **Figure 23**).
13) Click **Continue**.
14) Click **OK** or **Paste**.

Figure 23 Coding for the dummy-coded category *Doctor Rx - Yes*



What the Dummy Coded Variables Look Like in the SPSS Database. **Figure 24** presents the **Data View** portion of the SPSS database, describing the three new dummy coded variables that correspond to the original variable *Doctor Prescription Status*. For example, the first dummy coded variable *prescribedno*, indicates the number 1 (=yes) when the original variable *doctorprescribed* is 1 (=*Doctor Rx - No*). Otherwise there is a number 0 (=No). This same pattern is evident for the other values (*Doctor Rx - Recommended, Doctor Rx - Yes*) within the original *Doctor Prescription Status (i.e., doctorprescribed)* variable relative to the new dummy coded variables.

Figure 24 The dummy-coded variable *Doctor Prescription Status* in **Data View**

doctorprescribed	prescribedno	prescribedrec	prescribedyes
1.00	1.00	0.00	0.00
1.00	1.00	0.00	0.00
2.00	0.00	1.00	0.00
2.00	0.00	1.00	0.00
3.00	0.00	0.00	1.00
3.00	0.00	0.00	1.00

1= **Doctor Rx - No** in the original variable *Doctor Prescription Status*

1=**Yes** in the new dummy coded variable *Doctor Rx - No*

Please note the *values label* icon . Instead of pretending numbers are words, such as a *1* is a *Yes* and a *0* is a *No*, you can click this icon and it will show the corresponding world to the numbers in the SPSS database. For example, if you click the *values icon*, **Figure 24** will become the image in **Figure 25**. It may be easier to examine the cells in this form.

Figure 25 The dummy-coded variable *Doctor Prescription Status* in text response

doctorprescribed	prescribedno	prescribedrec	prescribedyes
No	Yes	No	No
No	Yes	No	No
Recommended	No	Yes	No
Recommended	No	Yes	No
Yes	No	No	Yes
Yes	No	No	Yes

1= **Doctor Rx - No** in the original variable *Doctor Prescription Status*

1=**Yes** in the new dummy coded variable *Doctor Rx - No*

After Dummy Coding You Must Select a Reference Group. Once the values (e.g., *Doctor Rx - No, Doctor Rx - Recommended, and Doctor Rx - Yes*) within the original categorical variable (e.g., *Doctor Prescription Status*) have been dummy coded, you must specify which value will be used as the **Reference Group** in the multiple linear regression model.

The **Reference Group** is the group to whom the other dependent variable scores will be compared. For example, if the group *Doctor Rx - No* within the variable *Doctor Prescription Status* is made the **Reference Group** in our study of levels of *Class Attendance* (Dependent Variable), then the analysis will tell us: Do *Class Attendance* Scores for those with a *Doctor Rx - Recommended* and/or *Doctor Rx - Yes* differ significantly **in reference to** those with a *Doctor Rx - No*? Thus, essentially, the **reference group** is the group the other groups are **compared to (or referenced)** regarding the scores on the dependent variable.

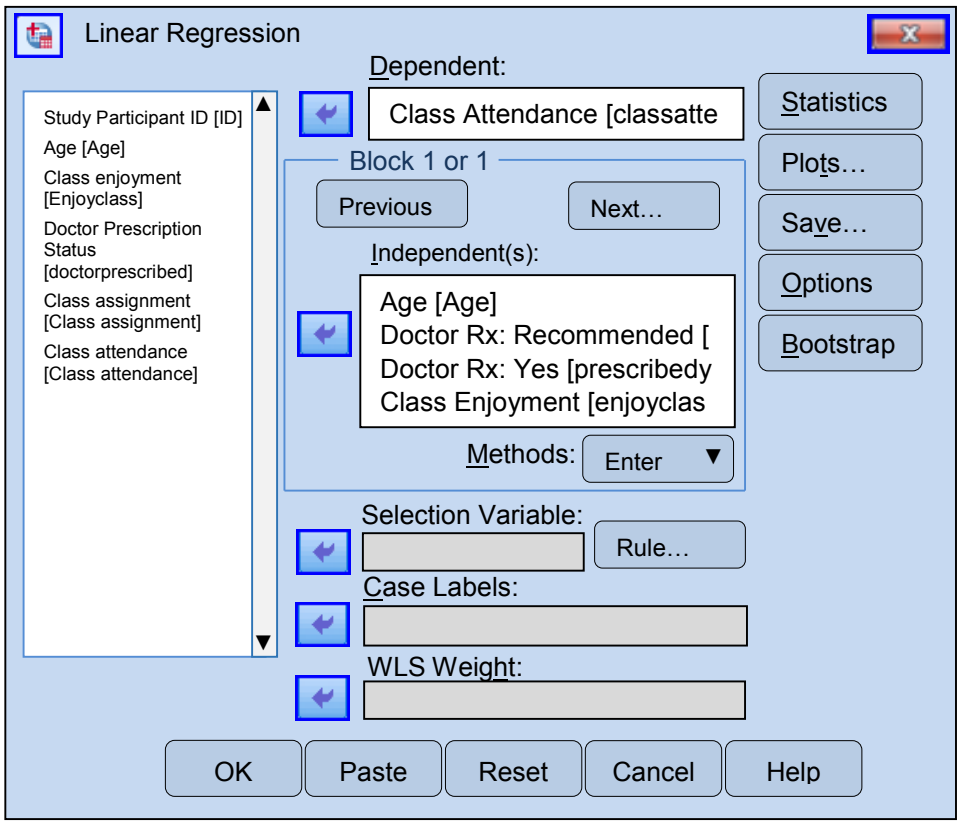
Selecting the Reference Group. There are several reasons for selecting one group as the reference group instead of the others. The first consideration is making the group of interest or under study the reference group. For example, if our study is focused on the well-being of individuals that have not been prescribed class by their doctor (*Doctor Rx - No*) then it would make sense to make that group the reference group. Then we would see the scores of *Doctor Rx - No* and *Doctor Rx - Recommended and Doctor Rx - Yes* relative to the group of interest *Doctor Rx - No*. The dummy-coded variable that represents the reference group (e.g., *Doctor Rx - No*) is left out of the multiple linear regression model, which will tell the SPSS software that this is the reference group.

Conducting Multiple Linear Regression Analysis

To conduct the multiple linear regression analysis procedure in SPSS go to:

1) **Analyze→ Regression→ Linear**.
2) In the dialogue box that opens, move the dependent variable *Class Attendance* from the left hand column into the box labeled **Dependent** on the right.

Figure 26 Multiple linear regression analysis: Predicting *Class Attendance*



3) Enter the predictor variables from the left hand column into the right hand column, labeled **Independent(s),** including: a) the dummy-coded covariate variables *Doctor Rx - Recommended* and *Education Doctor Rx - Yes* (leave out reference group *Doctor Rx - No*); The covariate variables *Age* and *Class Enjoyment*, as well as the independent variable *Class Assignment* (as presented in **Figure 26**).
4) Click **OK** or **Paste**.

Interpreting the Statistical Output: Multiple Linear Regression Analysis. **SPSS Output 7** presents the statistical output for the **Multiple Linear Regression Analysis** procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Class Attendance* and predictor variables (covariate and independent variables) in the context of the full model. Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

Overall Model. Note the second box in the output labeled **ANOVA**. In the final column labeled **Sig.**, the level of probability (statistical significance) for the overall test is indicated, which is **.000**. Thus, the overall model for the **multiple linear regression model** procedure is statistically significant ($p<.05$). Therefore, we will continue to examine the results of the test.

Individual Predictor Variables. Next, refer to the final table labeled **Coefficients**. This table provides information regarding the relationship between each predictor variable with the dependent variable. First, refer to the final column labeled **Sig.**, which presents the probability levels for these relationships. We see that all values are below .05 ($p<.05$), except for the covariate variables *Age* and *Doctor Rx - Recommended*. Subsequently, these finding indicate all the predictor variables have a statistically significant relationship with the dependent variable, except for *Age* and *Doctor Rx - Recommended*. **Thus, because the variables *Age* and *Doctor Rx - Recommended* have p-values greater that .05 ($p>.05$) we will not consider the relationship between those predictors and the dependent variable beyond this point**.

Part B Relationship Dimension 2: Directionality

The Unstandardized Beta (regression coefficient). Within the **SPSS Output 7**, refer to the final table labeled **Coefficients**. Refer to the first column under the label Unstandardized Coefficients labeled **B**, which indicates the directionality of the relationships between the predictor variables with the dependent variable. Specifically, this column presents the Unstandardized Beta. For example, observe the dummy-coded variable in the first column labeled, *Doctor Prescription Status Doctor Rx - Yes*. Notice that the unstandardized beta for each of these variables does not begin with a minus sign (1.425) indicating a positive relationship. However, recall there is a Reference group (*Doctor Prescription Status: Doctor Rx - No*) for these groups.

Therefore, the positive beta, along with the values reflecting statistical significance (p<.05 as Sig. = .000), indicates that study participants whose doctor prescribed meditation classes (*Doctor Rx - Yes*) attended classes a significantly higher number of days per week, i.e., values reflecting *Class Attendance* **in reference to** study participants whose doctors did not prescribe meditation class (*Doctor Rx - No*).

Next, refer to the next covariate predictor in this column, the variable *Class Enjoyment*. Again, we see under the Unstandardized Coefficients column, that the value for the beta (**B**) is positive (0.394) as there is not a minus sign, indicating a positive association. Specifically, this positive unstandardized beta indicates that as the covariate predictor variable *Class Enjoyment* increases, the dependent variable *Class Attendance* also increases.

Lastly, notice that the beta (**B**) for the independent variable *Class Assignment* is also positive as there is no minus sign in front of the unstandardized beta (1.477). In order to interpret the meaning of the negative unstandardized beta, recall that living with a *Morning* is coded as 1 and *Evening* is coded as 2 in the SPSS database. Thus, the positive beta suggests as the predictor *Class Assignment* increases from 1 to 2, or in conceptual terms from a *Morning* to *Evening*, *Class Attendance* scores go up (no minus indicates a positive relationship where scores increase together). Thus, assignment to evening class is associated with a significant increase in *Class Attendance*.

More About the Unstandardized Beta. It is important to note that the unstandardized beta also serves another purpose beyond indicating the directionality of a relationship between variables. Specifically, **the unstandardized beta also reflects the degree of change in the dependent variable relative to changes in the predictor variable**. Although a full discussion of the mechanics of the model in this respect is a bit beyond the focus of this text, this is an important consideration and fact to know. As a brief presentation, consider the classic model equation expressing this modeled change presented in **Figure 27**. In this model:

Y represents the **dependent variable**

X reflects a **predictor variable** (**independent** or **covariate variable**) response categories

b reflects the actual value of the **unstandardized beta** in the statistical output

Figure 27 The classic multiple linear regression model function equation

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

Thus, because we know our dependent variable, predictor variables (and response categories), and the unstandardized betas (as per the SPSS output), we can plug these numbers in and fill in the model in **Figure 28**. However, to keep these simple, let's just examine the *rate of change between one predictor variable, the independent variable (Class Assignment) and the dependent variable (Class Attendance)*. This equation is presented in **Figure 28**, where we see:

1) The first arrow indicates the **Y** represents the **dependent variable** *Class Attendance* (number of days attending class per week, 1-7).

2) The second arrow indicates the **b** reflects the value of **unstandardized beta** as per the SPSS statistical output (which is **1.477)** for the independent variable *Class Assignment*

3) Finally, the third arrow indicates the **X** represents the coding of the response categories within the **independent variable, which are 1 and 2,** where for the independent variable *Class Assignment,* Morning=1 and Evening=2.

Figure 28 Linear regression model function equation with only one predictor

$$Y = b_1X_1$$

Class Attendance = (1.477) (X = '1' or '2')

The change in the dependent variable (*Class Attendance* = the number of days attending class per week) occurs according to the response given for the independent variable. For example, if a study participant indicates he or she was assigned a **morning class** (coded with a **value of 1**) then the equation becomes: ***Class Attendance*** = (1.477) (1), which equals 1.477.

However, if a study participant indicates he or she was assigned an **evening class** (coded with a **value of 2**) then the equation becomes: ***Class Attendance*** = (1.477) (2), which equals 2.954. Therefore, is a study participants is assigned to the *evening class* he or she attends classes about one and a half (i.e., 1.477) more days per week (on average) relative to those assigned to the morning class.

SPSS Output 7

Regression

(3rd) Magnitude:
Large R^2 effect size

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.633	.400	.369	1.4436

(1st) Significance (model):
Statistically significant (p<.05)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.727	5	25.923	12.56	.000
	Residual	21.757	94	2.065		
	Total	56.484	99			

(1st) Significance (predictors):
Statistically significant (p<.05)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta (β)		
1	(Constant)	-.12	2.213		-.054	.957
	Age	-.002	.023	-.010	-0.9	.929
	Doctor Rx - Rec	.492	.353	.125	1.4	.166
	Doctor Rx – Yes	1.425	.351	.363	4.1	.000
	Class Enjoyment	.394	.141	.305	2.8	.006
	Class Assignment	1.477	.288	.410	5.1	.000

(2nd) Directionality:
Unstandardized Beta (B) indicates all positive relationships except for the IV

(3rd) Magnitude:
Standardized Beta (β) indicates the strongest predictor

Part C Relationship Dimension 3: Magnitude

Within the multiple linear regression model there are several means of indicating an effect size between the predictor and dependent variables. For example, the multiple linear regression analysis produces the standardized beta which indicates the strongest predictor within the regression model. In the regression model for sample study one, the highest standardized beta is .410 for the variable *Class Assignment*, which suggest this is the strongest predictor in the multiple linear regression model. However, the standardized beta is not the traditional effect size we might like to use to estimate the effect of the predictor variable on the dependent variable.

The multiple linear regression model also produces the R^2 effect size statistic that reflects the amount of variance in the dependent variable explained by the predictor variable. The **R^2** is .400, which as per the **SPSS Output 7** is a large effect size. It is also important note the **adjusted R^2**, which is a modified version of R^2 statistic. Specifically, the adjusted R^2 estimates how much variance in the dependent variable would be explained by the predictor variable(s) if the analysis was based upon the entire population from which the sample was derived. Generally, the closer the **R^2** and **adjusted R^2**, the more generalizable and credible is the estimate of the **R^2**.

For example, the **R^2** and **adjusted R^2**, in **SPSS Output 7** are quite similar (.400 and .369, respectively), which builds confidence in the generalizability of the findings. However, this effect size reflects the magnitude of the relationship between all the predictors and the dependent variable. Thus, we need another method to estimate the magnitude of the relationship between each individual predictor and the dependent variable.

However, the linear regression analysis procedure in SPSS does not have an option to estimate the effect size between variables. However, an equivalent procedure will produce the effect size estimate of the partial Eta squared statistic through the ANCOVA (Analysis of Covariance) statistical procedure. To perform an ANCOVA in SPSS go to:

1) **Analyze→ General Linear Models→ Univariate**.
2) Within the dialogue box that opens, move the dependent variable *Class Attendance* from the left hand column into the top cell under the term **Dependent Variable**.

3) Enter the categorical independent variable *Class Assignment* from the left hand column into the cell under the term **Fixed Factor(s)**.

4) Move the covariate predictor variables *Age, Doctor Prescription Status,* and *Class Enjoyment* from the left hand column into the cell under the term **Covariate(s)** (as presented in **Figure 28**).

5) Realize here we are treating the covariate variable *Doctor Prescription Status* as a continuous covariate predictor variable in this model. This is done for several reasons, including the fact that within an ANCOVA model, the covariates should be continuous in nature. However, as noted earlier, it is often acceptable to treat ordinal variables (ranked categorical variables) as continuous variables.

6) Click **Options**.

7) Within the dialogue box that opens, click the box for **Estimates of Effect Size** and **Descriptive Statistics** (as presented in **Figure 29**).

8) Click **Continue**.

9) Click **OK** or **Paste**.

Figure 28 ANCOVA: Producing an effect size for predictors of *Class Attendance*

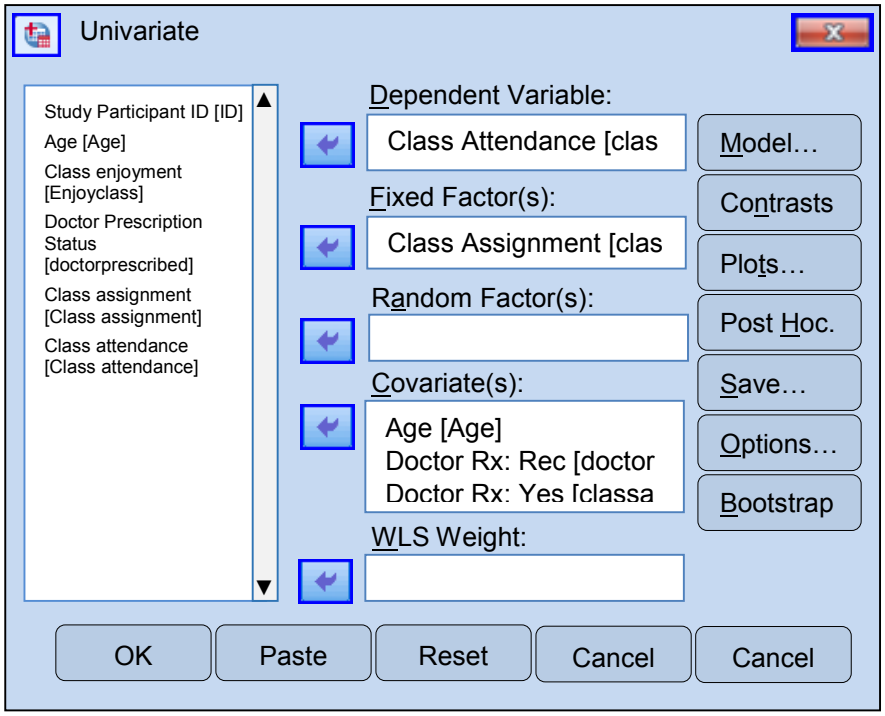
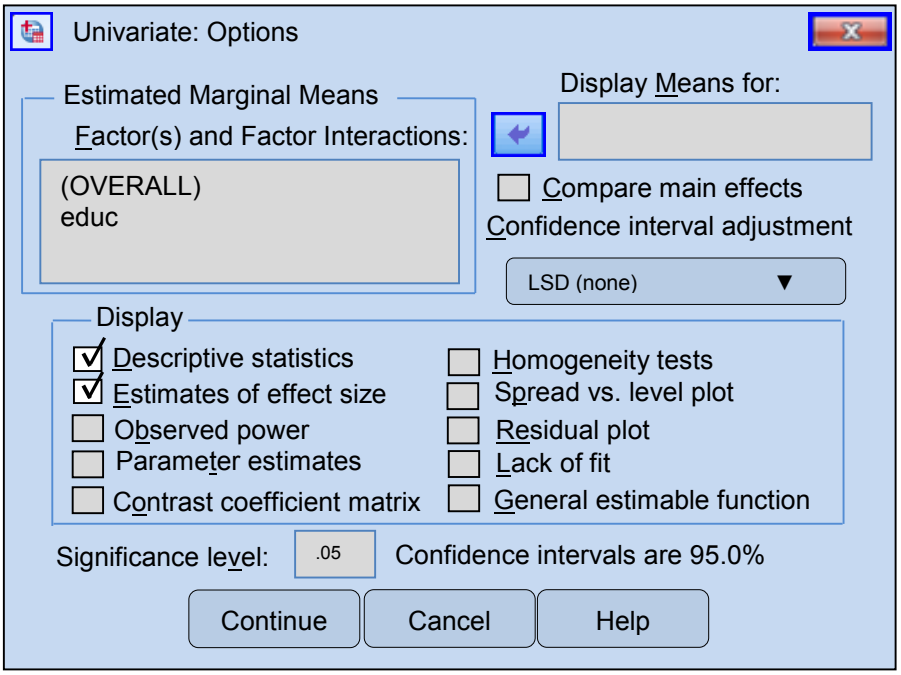


Figure 29 Activating **Estimates of Effect Size** and **Descriptive Statistics**



SPSS Output 8

Tests of Between-Subjects Effects

Dependent Variable: Class Attendance

Source	Type III Sum of Squares	df	Mean Sqr	F	Sig.	Partial Eta Squared
Corrected Model	128.62	4	32.15	15.659	.000	.397
Intercept	1.569	1	1.57	.764	.384	.008
Age	1.293	1	0.00	.000	.993	.000
doctorprescribed	33.509	1	33.51	16.32	.000	.147
enjoyclass	16.837	1	16.84	8.20	.005	.078
classassigned	54.938	1	54.94	26.76	.000	.220
Error	195.07	95	2.05			
Total	1683.09	100				
Corrected Total	323.693	99				

(3rd) Magnitude:

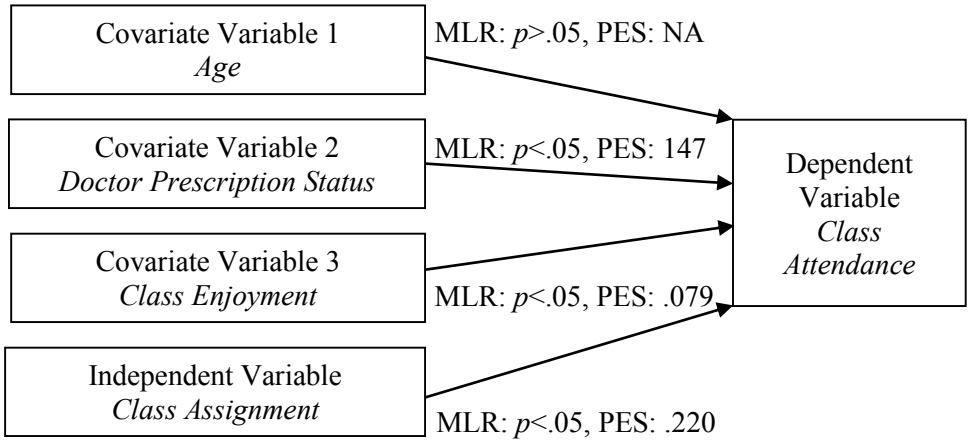
IV: Large effect size (PES = .22) (*p*<.05)

SPSS Output 8 presents statistical output for the ANCOVA procedure, which includes the partial Eta squared effect size. Specifically, the last column in the output labeled *Partial Eta Squared* presents the effect size estimates for the predictor/dependent variable relationships. Recall, the effect sizes for the partial Eta squared are small (>0.01), medium (>0.06), and large (>0.14). Thus, *Class Assignment* has a large effect size (0.22), which is larger in relation to the covariate variables. Therefore, we can state that according to these effect size estimates *Class Assignment* is the strongest predictor of *Class Attendance*.

What Did Multivariate Analysis Tell Us?

Figure 30 reflects the results on the multiple linear regression analysis used in sample study one examining which predictor variable is the strongest predictor of the dependent variable *Class Attendance*. We saw that at the multivariate level, all

Figure 30 What multivariate analysis told us in examining *Class Attendance*



predictors evidenced a statistically significant relationship (p<.05) with the dependent variable, with the exception of *Age* and *Doctor Prescription Status - Recommended*. Among all the variables that evidenced a statistically significant relationship with the dependent variable, *Class Assignment* had the largest effect size (0.22). Therefore, within the context of the full multivariate linear regression model, the strongest predictor of the dependent variable *Class Attendance* was the independent variable *Class Assignment*, while controlling for the influence of all covariate variables upon the dependent variable. In end, the response to the research question (**Is assigned class time the strongest predictor of the number of days class is attended per week?**), would be a resounding **Yes**.

Examining the Dependent Variable *Class Attendance* as a categorical variable:

Weekly Class Attendance 3 Days or Greater (Yes)

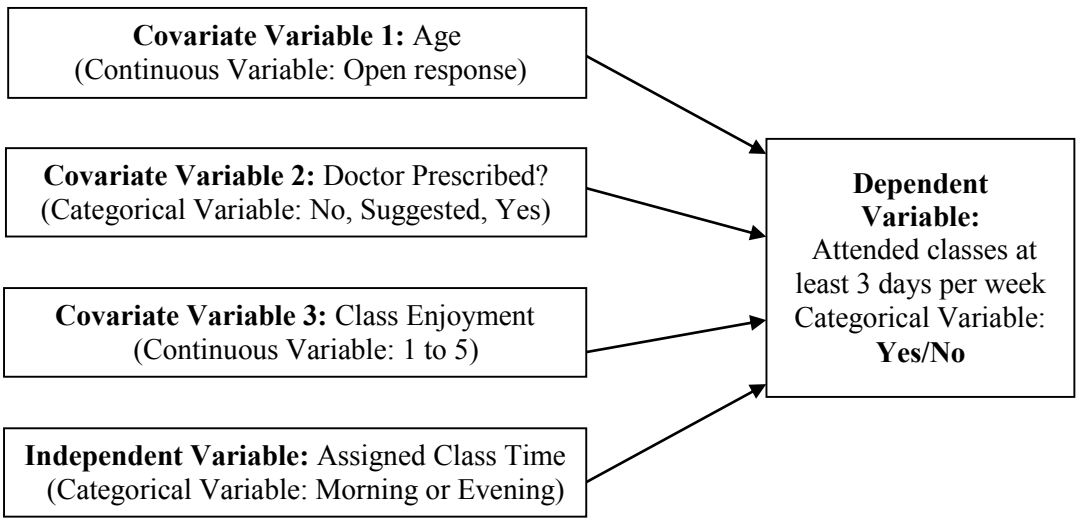
VS

Weekly Class Attendance Less than 3 Days (No)

Recall, the administrators of the retirement community also wanted us to examine the dependent variable *Class Attendance*, as a dichotomous categorical variable as Weekly Class Attendance 3 Days or Greater (coded as: Yes) VS Weekly Class Attendance Less than 3 Days (coded as: No). Recall, the research question posed is: **Is assigned class time the strongest predictor of attending class at least 3 days per week?**

The study map for this analysis is presented in **Figure 31**.

Figure 31 What multivariate analysis told us in examining *Class Attendance*

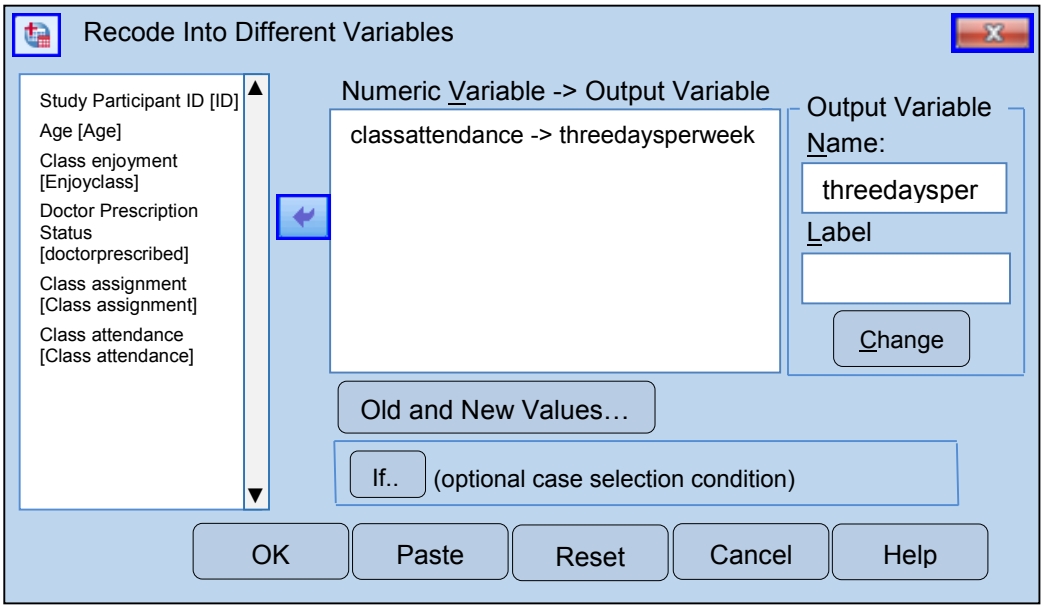


Recoding Class Attendance Into a Dichotomous Variable

In order to examine the dependent variable *Class Attendance* as a dichotomous variable, we must recode the original variable into the two categories *Attended Class 3 or More Days per Week* (Attended Class 3-7 Days per Week) and *Attended Class Less than 3 Days per Week* (Attended Class less than 3 Days per Week). To perform this recode in SPSS go to:

1) **Transform→ Recode Into Different Variables→**.
2) Use the center arrow to move the variable to be recoded, *Class Attendance*, from the left hand column into the right hand column.
3) In the box to the right labeled **Name**, enter the name of the new transformed variable, which will be *threedaysperweek*.
4) Click **Change**, which will move the new variable name into the center column, as presented in **Figure 32**.

Figure 32 Recoding the dependent variable *Class Attendance*



5) Click **Old and New Values**.

6) Recall, the full range of scores for *Class Attendance* is 1.00-7.00 and the cutoff score is above or below 3.00. To dichotomize the scale, we must code from the minimum score to the maximum score of the *Class Attendance* scale, while incorporating the cutoff score of 3.00. Therefore, scores reflecting *Attending Class Less than 3 Days per Week* (No) will be 0.00-2.99, while scores reflecting *Attending Class 3 or More Days per Week* (Yes) will be 3.00-7.00.

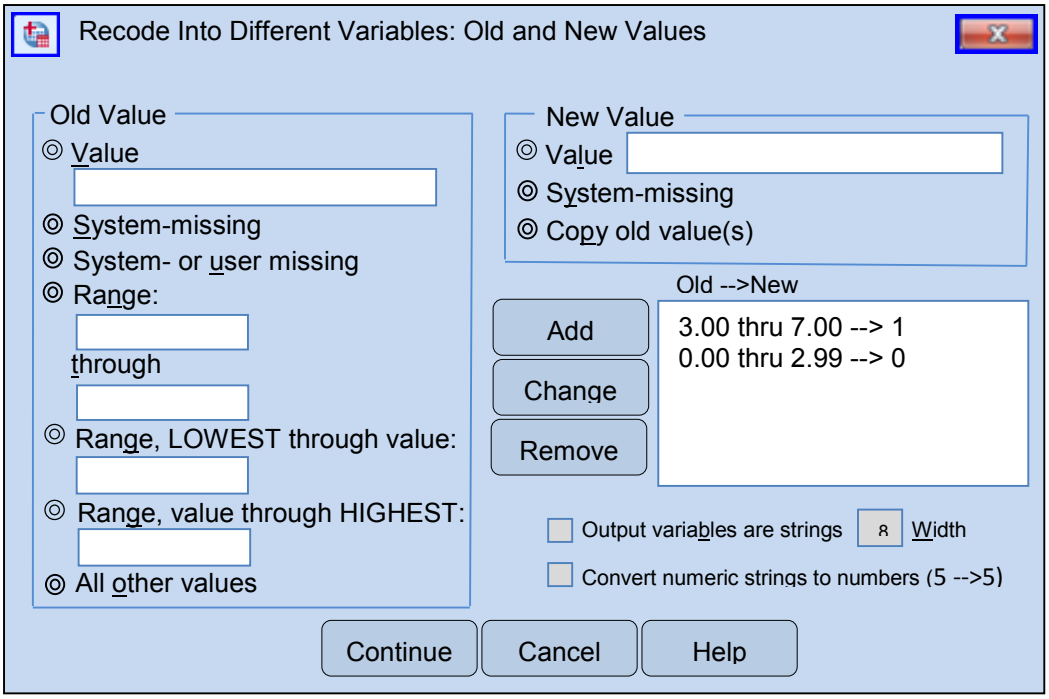
7) In the dialogue box that opens, within the left hand column labeled **Old Value**, click **Range**.

8) Enter the range of values reflecting *Attending Class Less than 3 Days per Week* (No), which is 0.00-2.99.

9) Within the right hand column labeled **New Value**, click **Value**, and enter a '0' (0=No) in the cell, click **Add**.

10) Within the left hand column labeled **Old Value**, click **Range**.

Figure 33 Recoding the dependent variable *Class Attendance* into *Happy* (Yes or No)



11) Enter the range of values reflecting *Attending Class 3 or More Days per Week* (Yes), which is 3.00-7.00.

12) Within the right hand column labeled **New Value**, click **Value** and enter a '1' (1=Yes) in the cell, click **Add** (as presented in **Figure 33**).

13) Click **Continue**.

14) Click **OK** or **Paste**.

15) Next, go into the **Variable View** portion of the SPSS database where you should find the newly created variable at the end of the list of variables.

16) Under **Label**, enter a description for the new dependent variable, such as *Attended Class at least 3 days per week*.

17) Under **Values**, enter a '1' for Yes and a '0' for No.

18) Check to be sure the recode was successful through conducting the frequencies procedure to examine the new dependent variable *threedaysperweek*.in relation to the original variable *classattendance*. Recall, to conduct the frequencies procedure in SPSS go to:

    a. **Analyze→ Descriptive Statistics→ Frequencies**.

    b. In the dialogue box that opens move the variable *threedaysperweek* from the left hand column into the right hand column.

    c. Click **OK** or **Paste**.

Instead of a continuous score distribution you should see values similar to **SPSS Output 9**. The output describes there are 29 (29%) scores below the cutoff point of 3 Days per Week and 71 (71%) cases above the cutoff point.

SPSS Output 9

Attended Class at least 3 days per week

		N	Percent	Valid Percent	Cumulative Percent
Valid	No	29	29.0	29.0	29.0
	Yes	71	71.0	71.0	100.0
	Total	100	100.0	100.0	

To check the accuracy of this recode, we can observe the frequency distribution of *Class Attendance* scores in their continuous form. To do this, conduct the frequency procedure using the variable *Class Attendance* in SPSS through going to:

a. **Analyze→ Descriptive Statistics→ Frequencies**.

b. In the dialogue box that opens move the variable *Class Attendance* from the left hand column into the right hand column.

c. Click **OK** or **Paste**.

SPSS Output 10 presents the frequency table, which will be included in the results of the frequencies procedure. We can check our recode by observing if within the original distribution, there is a 29/71 study participant split above and below the cutoff score of 3.00. Within the output, the location of the cutoff score of 3.00 is indicated using a dotted line. We can see there are 29 (29%) scores above the dotted line (see the *Cumulative Percent* column), which reflect the scores categorized less than 3 days per week, as well as 71 (71%) scores below the dotted line, which reflect the scores categorized as 3 or more days per week. Thus, as the recoded dichotomous dependent variable also identifies a group of 29 (29%) and 71 (71%), the recode seems successful.

SPSS Output 10

Class Attendance (Continuous Scores)

		Frequency	Percent	Valid Percent	Cumulative Percent
	0.00	2	2.0	2.0	2.0
	0.50	2	2.0	2.0	4.0
	0.80	2	2.0	2.0	6.0
	1.00	6	6.0	6.0	12.0
	1.20	1	1.0	1.0	13.0
> 3 Days (No)	1.80	3	3.0	3.0	16.0
	2.00	8	8.0	8.0	24.0
↑	2.20	4	4.0	4.0	28.0
3.00	2.90	1	1.0	1.0	29.0
Cutoff	3.00	12	12.0	12.0	41.0
Score	3.10	3	3.0	3.0	44.0
↓	3.20	1	1.0	1.0	45.0
≤3 Days (Yes)					
	Total	100	100.0	100.0	

STEP 5: BIVARIATE ANALYSIS

Figure 34 presents The Bivariate Test Key again, but this time with the study variables plugged in for the analysis of the dependent variable measured as attending class 3 days a week or more vs. less than 3 days per week. The key instructs to use an independent-samples t-test to examine the relationship between the dependent variable *Attended Class at Least 3 Days per Week* and covariate variables *Age* and *Class Enjoyment*, is an Independent Samples T-Test. Furthermore, a chi-square analysis would be used to relate the other predictors (*Doctor Prescription Status* and *Class Assignment*) with the outcome variable. However, since the Independent Samples T-Test was illustrated in the last section, for the sake of brevity we will only present the chi-square test relating the independent variable *Class Assignment* and the dependent variable here. However, all covariate variables will be included in the regression model in this analysis.

Figure 34 The Bivariate Test Key (Study Variables Plugged In)

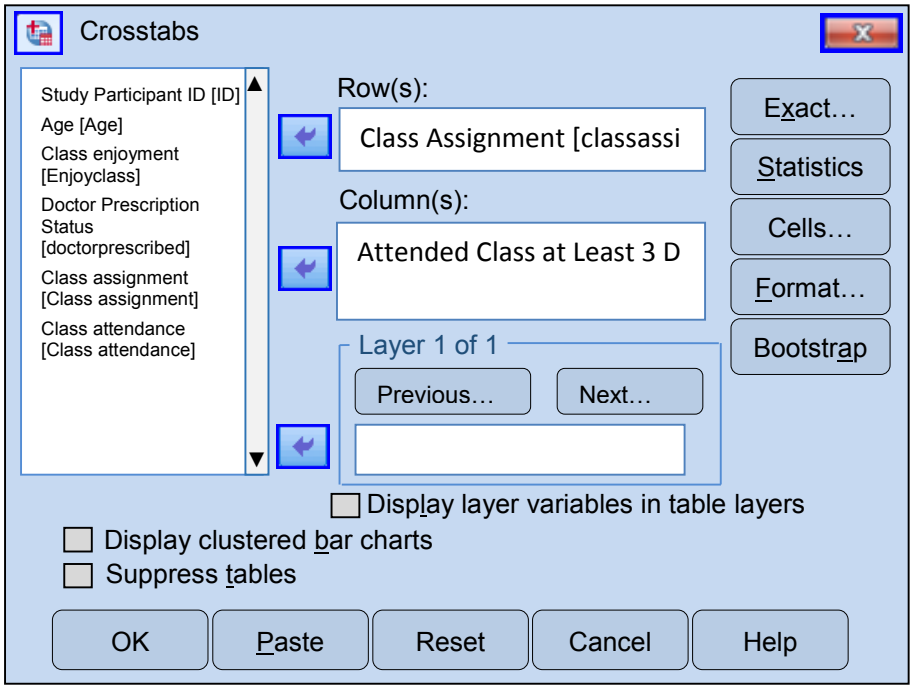
Variable 1	Variable 2	Test To Use
Categorical (2 categories)	Continuous	Independent Samples T-Test
Attend 3 Days (Yes, No)	Age (NA)	Independent Samples T-Test
Attend 3 Days (Yes, No)	Class Enjoyment (1-5)	Independent Samples T-Test
Categorical (≥3 categories)	Continuous	One-Way ANOVA
Continuous	Continuous	Correlation
Categorical	Categorical	Chi-Square
Attend 3 Days (Yes, No)	Doctor Rx (N/R/Y)	Chi-Square
Attend 3 Days (Yes, No)	Class Assignment (M, E)	Chi-Square

Chi-Square

The bivariate test key indicates the appropriate statistical procedure to examine the relationship between two categorical variables is a chi-square analysis. Therefore, we will use this procedure to examine the relationships between the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No) with the independent variable *Class Assignment*. The test assumptions for a chi-square include that each observation (e.g., study participant) is independent of all the others. Specifically, each study participant is within only one group, such as either in the group that is assigned to class in the Morning or Evening, but not both in one class session.

Additionally, there is an assumption for a chi-square table larger than 2x2, where no more than 20% of the expected counts (expected counts may be calculated by checking the *Expected* box under the *Cells* option of the chi-square test) are less than 5 and all individual counts are one or greater (Yates, Moore, and McCabe, 1999, p. 734). Furthermore, when a chi-square table is 2x2, there is an assumption that all expected counts are approximately 10 or greater (Cochran, 1952; 1954). When expected counts are somewhat small (<10), the *Fisher exact test for 2x2 contingency table*, might be considered. Once, these assumptions are considered, the chi-square analysis can be approached.

Figure 35 **Chi-Square** *Attended Class at Least 3 Days per Week* by *Class Assignment*



To conduct the chi-square using SPSS, go to:

1) **Analyze→ Descriptive Statistics→ Crosstabs**.
2) In the dialogue box that opens, use the arrow in the upper center to move the predictor variable (e.g., *Class Assignment*) into the cell on the right hand side marked **Row(s)**.
3) Move the dependent variable *Attended Class at Least 3 Days per Week* into the cell marked **Column(s)** as presented in **Figure 35**.

Figure 36 Activating the Chi-square statistic and Phi and Cramer's V

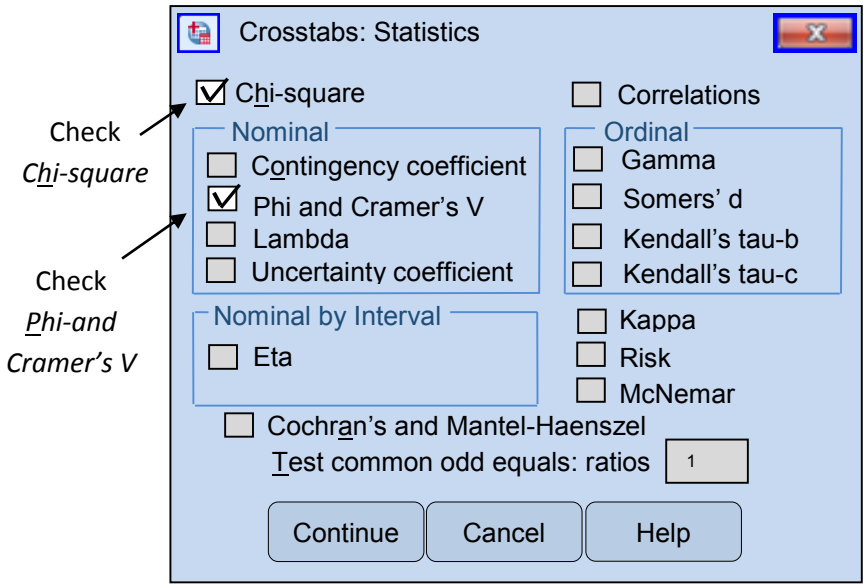
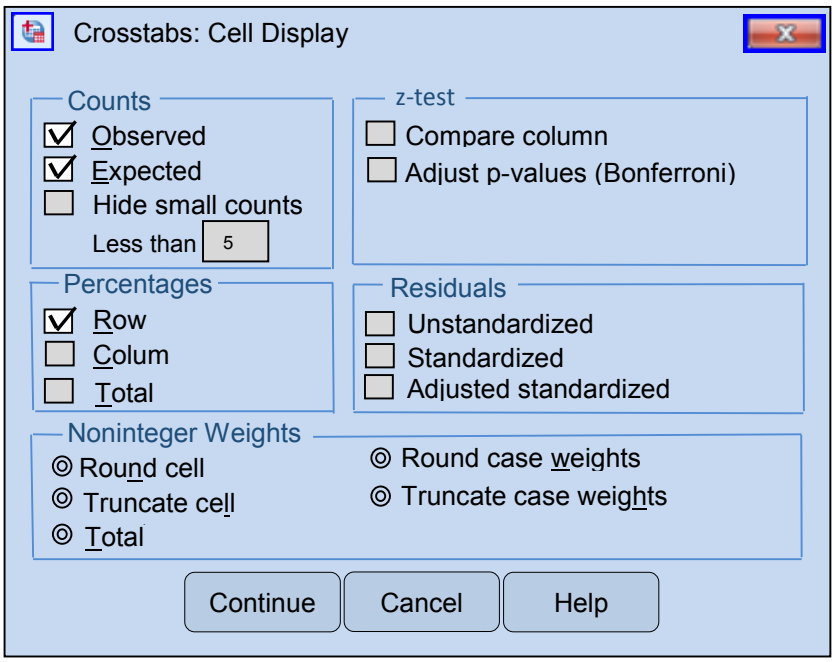


Figure 37 Activating the Observed and Expected Counts and Row Percentages



4) Click the button marked **Statistics**.
5) In the dialogue box that opens, check the boxes next to **Chi-Square** and **Phi and Cramer's V** as presented in **Figure 36**.
6) Click **Continue**.
7) Click the button marked **Cells**.
8) In the dialogue box that opens, under **Counts**, check **Observed** and **Expected**.
9) Under **Percentages**, check the box next to **Row** as presented in **Figure 37**.
10) Click **Continue**.
11) Click **OK** or **Paste**.

Interpreting the Statistical Output: Chi-Square. **SPSS Output 11** presents the statistical output for the chi-square procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No) and the independent variable *Class Assignment* (Morning or Evening). Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

To observe significance, look at the second box in the **SPSS Output 11** labeled **Chi-Square Tests**. Within this box, the first line reads **Pearson Chi-Square**. If the value to the far right on this line, under the column **Asymp. Sig. (2-sided)**, is .05 or below ($p<.05$), then the test indicates a statistically significant relationship between variables. Within **SPSS Output 11**, this cell presents **the value .001, which indicates that this test achieves statistical significance**.

Part B Relationship Dimension 2: Directionality

Next, refer to the first box labeled **Class Assignment*Attended Class at Least 3 Days per Week Crosstabulation**. Within this box, under the column labeled **Attended Class at Least 3 Days per Week**, note the column marked **Yes**. In the intersecting first row for the variable *Class Assignment* we see that 56.0% (28 of 50) of study participants assigned to morning classes are coded as **Attended Class at Least 3 Days per Week** (Yes). Additionally, we see that 86.0% (43 of 50) of study participants assigned to evening classes are coded as **Attended Class at Least 3 Days per Week** (Yes).

Regarding the **Count** relative to the **Expected Count**, we see there is an overrepresentation of residents in the evening classes and underrepresentation of residents in morning classes. Thus, we can clearly see that assignment to evening classes reflects a greater likelihood of attending classes at least 3 days per week relative to those assigned to morning classes.

SPSS Output 11

Crosstabs

Class Assignment* Attended Class at Least 3 Days per Week Crosstabulation

			Attended Class at Least 3 Days per Week		Total
			No	Yes	
Class Assignment	Morning	Count	22	28	50
		Expected Count	14.5	35.5	50.0
		% within classassignment	44.0%	56.0%	100.0%
	Evening	Count	7	43	50
		Expected Count	14.5	35.5	50.0
		% within classassignment	14.0%	86.0%	100.0%
Total		Count	29	71	100
		Expected Count	29.0	71.0	100.0
		% within classassignment	29.0%	71.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.928	1	.001
Continuity Correction	9.52		.002
Likelihood Ratio	11.3	1	.001
Fisher's Exact Test			
Linear-by-Linear	10.8	1	.001
N of Valid Cases	100		

(2nd) Directionality:
Assigned Evening = Higher % *Yes*

(1st) Significance:
Statistically significant (*p*<.05)

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.331	
	Cramer's V	.331	.001
	N of Valid Cases	100	.001

(3rd) Magnitude:
Medium effect size (Phi = 0.33)

Part C Relationship Dimension 3: Magnitude

To test the effect size for a chi-square analysis of variables presented in a 2 X 2 table, we can determine the value of Phi (φ). The value of **Phi** is presented in the third box within our **SPSS Output 11**. This estimate of Phi within the statistical output is .331. Recall, the effect sizes for Phi are 0.1 (small), 0.3 (medium) and large (0.5). Thus, our effect size of .3652 reflects a medium effect size, indicating that the independent variable *Class Assignment* has a medium size effect on the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No).

Summary. Overall, the chi-square analysis indicated a statistically significant relationship between the independent and dependent variables ($p<.05$). In terms of directionality, there seems to be an association between assignment to evening classes with a greater likelihood of *Attending Class at Least 3 Days per Week* (Yes or No), relative to those assigned to morning classes (86% vs. 56%, respectively). Lastly, in terms of magnitude, the the independent variable *Class Assignment* evidenced a medium size effect (Phi = 0.33) upon the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No).

Step 6: Multivariate Analysis

In this section, we will employ multivariate analysis to identify the strongest predictor the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No). However, prior to analysis the correct multivariate test must be selected. **Figure 38** presents the **Multivariate Test Key with the variables for plugged in for the current analysis**. This key describes the appropriate multivariate procedure when examining a dichotomous categorical dependent variable, such as *Attended Class at Least 3 Days per Week* (Yes or No), is a binary logistic regression. Binary logistic regression can be used to predict the value of a dichotomous dependent variable based on the value of two or more other predictor variables, while observing the relative contribution of each.

The main **test assumptions** that to be examined is a check for significant **multicollinearity**. Of course, categorical predictors with 3 or more categories, such as *Doctor Rx Status*, must also be dummy-coded for this regression model, just as we did earlier for the multiple regression model. Subsequently, these dummy-coded versions of this variable will be used in the binary regression model in this section.

Figure 38 Multivariate Test Key with the study variable plugged in

Dependent Variable	Regression to Use
Categorical (≥3 categories)	Multinomial Logistic (Not Presented here)
Categorical (2 categories)	Binary Logistic
Categorical (2 categories)	Binary Logistic
Continuous	Linear

Conducting Binary Logistic Regression

Our goal in multivariate testing is to test the association between multiple predictor variables with the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No). To conduct a Binary Regression model in SPSS go to:

1) **Analyze→ Regression→ Binary Logistic**.
2) In the dialogue box that opens, move the dependent variable *Attended Class at Least 3 Days per Week* (Yes or No) from the left hand column into the right hand column into the cell labeled **Dependent**.
3) Enter the covariate predictor variable *Age* into the cell labeled **Covariates**.
4) Enter the two dummy-coded variables *Doctor Rx - Recommended* and *Doctor Rx - Yes* (reflecting *Doctor Prescription Status*) into the cell labeled **Covariates**. Recall, we are leaving the third dummy-coded variable *Doctor Rx - No* out of the model to serve as the reference group.
5) Enter the covariate predictor variable *Class Enjoyment* into the cell labeled **Covariates**.
6) Enter the independent variable *Class Assignment* into the cell labeled **Covariates**, as presented in **Figure 39**.
7) Click the button labeled **Options**.
8) In the dialogue box that opens, under the section **Statistics and Plots**, check the box next to **CI for exp(B)** (as presented in **Figure 40**), which will produce the 95% confidence interval.
9) Click **Continue**.
10) Click **OK** or **Paste**.

Figure 39 Conducting binary logistic regression

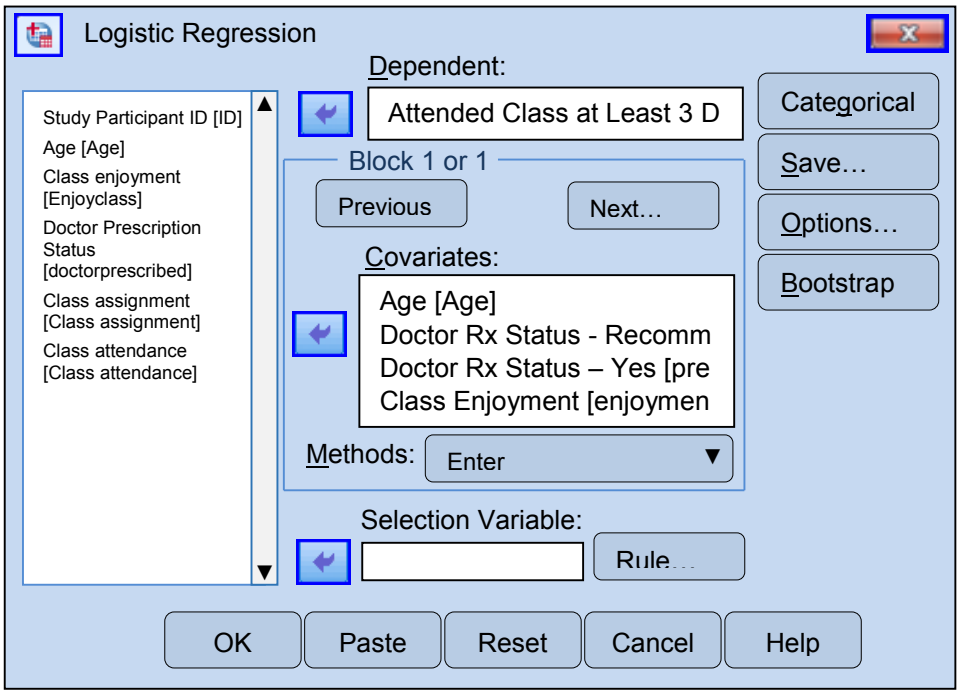
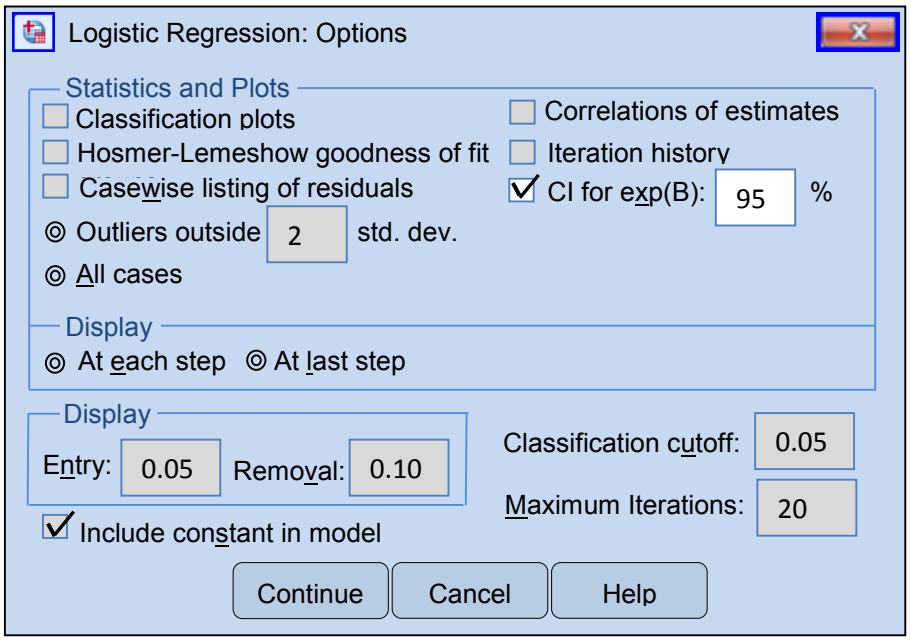


Figure 40 Activating the 95% confidence interval in binary logistic regression



Interpreting the Statistical Output: Binary Logistic Regression Analysis. **SPSS Output 12** presents the statistical output for the **Binary Logistic Regression Analysis** procedure. In this section we use this statistical output to identify the three dimensions of the relationship between the *Attended Class at Least 3 Days per Week* (Yes or No) and predictor variables (covariate and independent variables) in the context of the full model. Specifically, we will present the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

Part A Relationship Dimension 1: Significance

Significance of the Overall Model. Within **SPSS Output 12**, the first box (labeled **Omnibus Tests of Model Coefficients**) presents the level of statistical significance for the overall binary logistic regression model. Specifically, in the bottom line (labeled **Model**) we see the value of the chi-square statistic (27.931) and level of statistical significance (**Sig.**=.000) for the regression model. If the probability level is greater than .05, there is no need to evaluate the rest of the model as the findings are regarded as occurring by chance. However, we see in this output that the level of probability for the binary logistic regression model for sample study two is .000. Thus, we will review the model further. While assessing the full model, you might also note the percentage of cases categorized correctly, which is 82%. This is another common metric reported for the logistic regression model.

Significance of the Individual Predictor Variables. Next, within the box labeled **Variables in the Equation (A)**, note the last column labeled **Sig.**, which presents the level of statistical significance between each individual predictor and the dependent variable. The top two rows of this column describe the covariate variable *Age* and the dummy-coded variable *Doctor Rx - Recommended* have probability levels greater than .05 and do not have a statistically significant relationship with the dependent variable. The dummy-coded variable *Doctor Rx - Yes* does have a p-value below .05. This indicates that in reference to the reference group, which is *Doctor Rx - No,* the residents whose doctors did prescribe meditation class (*Doctor Rx - Yes)* are significantly more likely to attend meditation classes at least 3 days per week. The next two lines within the **Sig.** column reflects the relationship between the covariate variable *Class Enjoyment* and independent variable *Class Assignment* are related to the dependent variable at a statistically significant level.

Part B Relationship Dimension 2: Directionality

The Beta. Next, to identify directionality, again refer to the column labeled **B** within the table titled **Variables in the Equation (A)**, which presents the beta coefficient**.** Please note, all variables that are related to the dependent variable at a statistically significant level (*Doctor Rx - Yes*, *Class Enjoyment*, and *Class Assignment*) all have positive relationships with the dependent variable *as each beta coefficient does not have a minus sign*.

Part C Relationship Dimension 3: Magnitude

The effect size in binary logistic regression between relationships is expressed via the odds ratio. Although estimates of what constitutes a small, medium, and large effect size for an odds ratio often varies, a generally accepted convention is **small = 1.44, medium = 2.47, and large = 4.25**. Within the statistical output the odds ratio is reported in the column labeled **Exp(B)**. The confidence interval for each odds ratio is presented in the two columns (lower and upper bound estimates) to the right.

The 95% confidence interval. The 95% confidence interval is our way of acknowledging that the odds ratio generated in statistical analysis is not likely the actual odds ratio that exists in the target population as a whole. However, we are suggesting that we are 95% confident that the true odds ratio within the target population exists between the values presented in the 95% confidence interval. The wider the 95% confidence interval, the less stability is indicated in the finding.

The odds ratios for this study can be found within the box labeled **Variables in Equation (B)** in the column marked **Exp(B)**. Within this column, on the third line, we see the odd ratio for the dummy-coded variable *Doctor - Rx: Yes* is 4.72. This value suggests that study participants in the group *Doctor - Rx: Yes* are 4.72 times more likely to be *Attend Class 3 or More Days per Week* (Yes), in reference to study participants in the *Doctor - Rx: No* group (the reference group). Thus, a great magnitude is indicated as there is a large effect size between the variable *Doctor Prescription Status* and *Attended Class 3 or More Days per Week* (Yes or No). However, the 95% confidence interval for this odds ratio is extremely wide (95% CI = 1.23-18.16), suggesting the finding is unstable.

Next, on the third line, within the column **Exp(B)**, we see an odds ratio of 1.74 for the covariate variable *Class Enjoyment.* However, it is important to realize that an odds ratio in a logistic regression model is *not interpretable for a continuous predictor variable in the same way as a categorical predictor variable, where one might state an*

SPSS Output 12

Logistic Regression

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	27.931	5	.000
	Block	27.931	5	.000
	Model	27.931	5	.000

(1st) Significance (model): Statistically significant (*p*<.05)

Classification Table[a]

Observed		Predicted		
		Is the study participant happy?		Percent Correct
		No	Yes	
Attended Class At Least 3 Day	No	16	13	55.2
	Yes	5	66	93.0
Overall percent				82.0

Percentage of cases categorized correctly

(1st) Significance (predictors): 3 Statistically significant (*p*<.05)

(2nd) Directionality: Beta (B) indicates all sig. relationships are positive

Variables in Equation (A)		B	S.E.	Wald	df	Sig.
Step 1	Age	-.02	.04	.26	1	.613
	Doctor Rx - Recommended	.77	.62	1.54	1	.214
	Doctor Rx - Yes	1.55	.69	5.09		.024
	Class Enjoyment	.56	.27	4.41	1	.036
	Class Assignment	1.86	.57	10.66	1	.001
	Constant	-2.17	4.06	.28	1	.594

(3rd) Magnitude: Odds ratio effect sizes are large

Variables in Equation (B)		Exp (B)	95% Confidence Interval of the Difference	
			Lower	Upper
Step 1	Age	.979	.90	1.06
	Doctor Rx - Recommended	2.150	.64	7.20
	Doctor Rx - Yes	4.720	1.23	18.16
	Class Enjoyment	1.744	1.04	4.93
	Class Assignment	6.442	2.11	19.72
	Constant	.115		

outcome is more or less likely to occur. Specifically, when the predictor is continuous, the odds ratio may be reported, but not phrased in terms of the event being more or less likely to happen. For example, in the case of the covariate predictor variable *Class Enjoyment*, we might choose to simply report that a higher *Class Enjoyment* is significantly associated with experiencing the event, but refrain from making a statement regarding how many times more or less likely a study participant is to experience the outcome.

The fourth line within the column labeled **Exp(B)**, presents the odds ratio between the independent variable *Class Assignment* and the dependent variable *Attended Class 3 or More Days per Week* (Yes or No), which is 6.44. Within the independent variable *Class Assignment*, morning is coded as 1 and evening is coded as 2. Thus, the positive relationship (e.g., no minus in front of the beta) suggests as the variable moved from 1 to 2, i.e., morning to evening, there is an almost 6 and a half times (OR=6.44) greater likelihood to experience the dependent variable *Attended Class 3 or More Days per Week* (Yes or No). The 95% confidence interval is a bit wide ranging from 2.11 to 19.72, which suggests that the findings are not exceptionally definitive.

What Did Multivariate Analysis Tell Us?

Figure 41 describes the findings within the multivariate analysis for this analysis. The model has indicated that while the covariate variable *Age* was not significantly associated with the dependent variable, while residents that were prescribed meditation classes by their doctor (i.e., *Doctor Rx - Yes*) were almost five times (OR=4.72) more likely to *Attend Classes 3 Days a Week or More* (Yes), relative to the residents whose doctor did not prescribe meditation classes (*Doctor Rx - No* group). The covariate variable *Class Enjoyment* was also significantly related to the dependent variable in the full regression model. Lastly, in terms of *Class Assignment*, residents assigned to evening class were almost six and a half (OR=6.44) more likely to attend class 3 or more days per week relative to those assigned class in the morning.

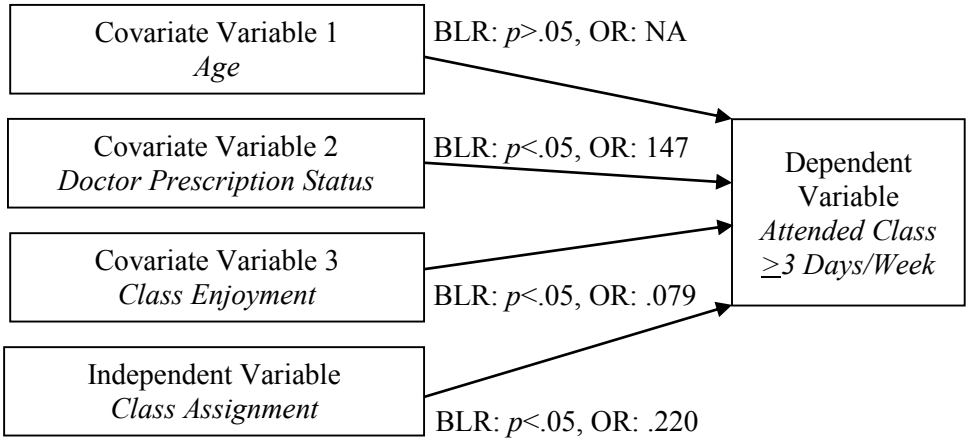



Figure 41 What multivariate analysis told us in examining *Class Attendance*