Week 5:
Multivariate Analysis

The Steps of Data Analysis

6 Week Webinar Class

Presented By

William Bannon, PhD

StatsWhisperer

New York, New York

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 (*Education Level*, *Income Level*, *Do You Live with a Dog or a Cat?*) is the strongest predictor of the dependent variable *Happiness*, which was the overall goal of our data analysis study.

Page 204 of *The 7 Steps of Data Analysis* (TSSODA) textbook presents **Figure 4.105 The Multivariate Test Key (also entered on the next page)**, which describes that selecting the appropriate multivariate statistical procedure depends upon the structure (i.e., categorical or continuous) of the dependent variable. Furthermore, **Figure 4.106 (also on page 204 of TSSODA, as well as on the next page)** presents the **Multivariate Test Key with the study variable plugged in**. Specifically, we see the key with the dependent variable for sample study 1 plugged in. The key instructs that the appropriate regression model to use when examining a continuous dependent variable, such as *Happiness* (Response Category: 1-5) 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 can be used to predict the value of the continuous dependent variable based on the value of two or more other predictor variables. Multiple linear regression will also allow us to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained in the dependent variable. We know these data are appropriate for this procedure as in **Step 3: Checks of Data Integrity**, we assessed if several assumptions of multiple linear regression analysis were met including **linearity**, **normality**, **no undue influence of outlier scores**, **independence of observations, homoscedasticity**, as well as other necessary considerations.

Figure 4.105 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 4.106 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 (1-5)	Happiness (1-5)

Dummy-Coding Variables

Pages 204-212 of TSSODA also reviews the important recode function known as **Dummy-Coding**. The text describes that dummy coding is the key to using a categorical variable with three or more response categories in a 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. Please see these pages for a detailed description of this concept and process, where the variable *Education Level* is dummy-coded for use in a regression model as an example.

Sample Study 1: Conducting Multiple Linear Regression Analysis

Pages 212-219 present the process of conducting the multiple linear regression analysis procedure in SPSS. The interpretation of the statistical output (please see **SPSS Output 4.30** on page 216 of TSSODA, as well as on the next page) begins on page 213. In this section we use this statistical output to identify the three dimensions of the relationship between the dependent variable *Happiness* and predictor variables (covariate and independent variables) in the context of the full model. Specifically, we see the relationship dimensions of significance, directionality, and magnitude in three separate parts (Part A, Part B, and Part C, respectively).

SPSS Output 4.30

Regression

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

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.784	.615	.599	.47856

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

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.727	4	8.682	37.91	.000
	Residual	21.757	95	.229		
	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)	3.43	.183		18.7	.000
	Education Level: College	.382	.123	.252	3.7	.000
	Education Level: Graduate School	.698	.118	.333	5.6	.000
	Income Level	.010	.002	.386	5.7	.000
	Do You Live with a Dog or a Cat?	-.461	.097	-.306	-4.7	.000

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

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

Part A Relationship Dimension 1: Significance

Overall Model. Within **SPSS Output 4.30** please 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), which indicates that all the predictor variables have a statistically significant relationship with the dependent variable. If a value greater that .05 (p>.05) we would not consider the relationship between that predictor and the dependent variable.

Part B Relationship Dimension 2: Directionality

The Unstandardized Beta (regression coefficient). Within the **SPSS Output 4.30**, 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 two dummy-coded variables at the top of the first column, *Education Level: College* and *Education Level Graduate School*. Notice that the unstandardized beta for each of these variables does not begin with a minus sign (.382 and .698) indicating a positive relationship. However, recall there is a Reference group (*Education Level: High School/GED*) for these groups.

Therefore, the positive betas, along with the values reflecting statistical significance (p<.05 as Sig. = .000 and .000), indicate that study participants with a level of education at the *College* and *Graduate School* levels have significantly higher scores (recall positive relationship indicates higher scores) reflecting *Happiness* **in reference to** study participants with a *High School/GED* level of education.

Next, refer to covariate predictor variable *Income Level*. Again, we see under the Unstandardized Coefficients column, that the value for the beta (**B**) is positive (0.01) as there is not a minus sign, indicating a positive association. Specifically, this positive unstandardized beta indicates that as the covariate predictor variable *Income Level* increases, the dependent variable *Happiness* also increases. However, notice that the

beta (**B**) for the independent variable *Do You Live with a Dog or Cat?* is negative as there is a minus sign in front of the unstandardized beta.

In order to interpret the meaning of the negative unstandardized beta, recall that living with a dog is coded as 1 and a cat 2 in the SPSS database. Thus, the negative beta suggests as the predictor *Do You Live with a Dog or Cat?* increases from 1 to 2, or in conceptual terms from a *dog* to a *cat*, *Happiness* scores go down (the minus indicates a negative relationship where scores decrease). In other words, living with a cat is associated with significant decreases in *Happiness* scores.

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 4.116**. 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 4.116 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 4.117**. However, to keep these simple, let's just examine the *rate of change between one predictor variable, the independent variable (Do You Live with a Dog or a Cat?) and the dependent variable (Happiness)*. This equation is presented in **Figure 4.117**, where we see:

1) The first arrow indicates the **Y** represents the **dependent variable** *Happiness*.

2) The second arrow indicates the **b** reflects the value of **unstandardized beta** as per the SPSS statistical output (which is **-0.461)** for the independent variable *Do You Live with a Dog or a Cat?*

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 *Do You Live with a Dog or a Cat?,* Dog=1 and Cat=2.

Figure 4.117 Linear regression model function equation with only one predictor

$$Y = b_1X_1$$

Happiness = (-0.461) (X = '1' or '2')

The change in the dependent variable occurs according to the response given for the independent variable. For example, if a study participant indicates he or she **lives with a cat** (coded with a **value of 2**) then the equation becomes: ***Happiness*** = (-0.461) (2).

Therefore, for study participants that live with a cat, their *Happiness* scores decreases by .922 on average as: ***Happiness*** = (-0.461) (2), which is -.922.

If a study participant indicates he or she lives with a dog (coded with a value of 1) then the equation becomes: ***Happiness*** = (-0.461) (1).

Therefore, for study participants that live with a dog, their *Happiness* score decreases by .461 on average as: ***Happiness*** = (-0.461) (1), which is -.461.

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 .583 for the variable *Education Level: Graduate School*, 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. In sample study one the **R^2** is .615, which as per the **Figure 3.13** 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 4.30** are quite similar (.615 and .599, 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, which is presented on pages 217-219 of TSSODA text. **SPSS Output 4.31** (presented on page 219 of TSSODA, as well as below) presents the 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).

SPSS Output 4.31

Tests of Between-Subjects Effects

Dependent Variable: Happiness

Source	Type III Sum of Squares	df	Mean Sqr	F	Sig.	Partial Eta Squared
Corrected Model	46.100	3	15.37	142.06	.000	.816
Intercept	23.229	1	23.23	214.75	.000	.691
educ	1.293	1	1.29	11.95	.001	.111
income	18.923	1	18.92	174.94	.000	.646
livedogcat	1.057	1	1.06	9.78	.002	.092
Error	10.384	96	.11			
Total	803.960	100				
Corrected Total	56.484	99				

(3rd) Magnitude:

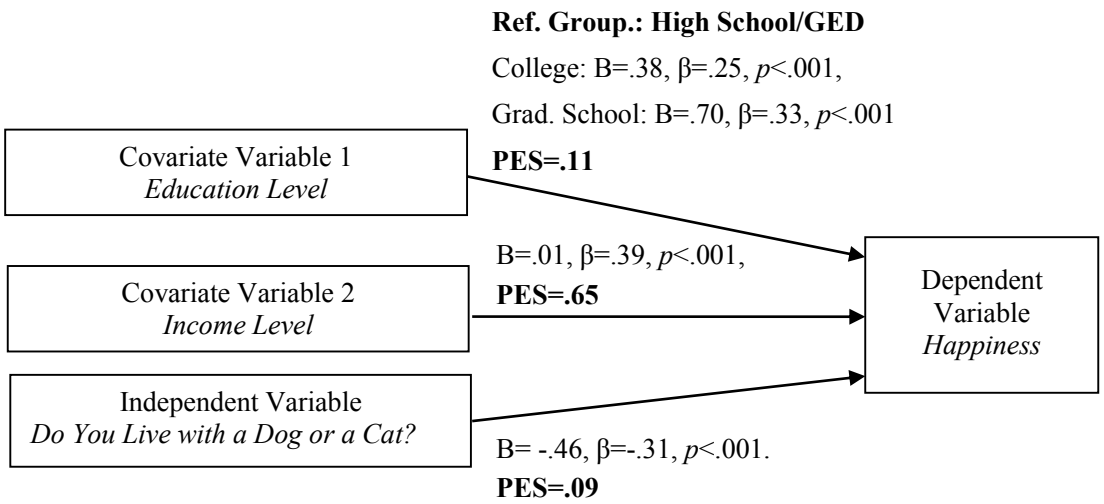
IV: Medium to Large effect size (PES = .092)

Thus, living with a dog or cat has a medium to large effect size (0.092), as does *Education Level* (0.111), while *Income Level* (0.646) has a very large effect size. These estimates reflect the strength of predictors described by the Standardized Beta in the multiple linear regression model where *Income Level* was the strongest predictor of *Happiness*.

What Did Multivariate Analysis Tell Us?

Figure 4.119 (presented on page 220 of SSODA, as well as below) 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 *Happiness*. We saw that at the multivariate level, all predictors evidenced a statistically significant relationship ($p<.05$) with the dependent variable. After the effects of both covariate predictor variables *Education Level* and *Income Level* were controlled for in the multivariate model, the independent variable *Do You Live with a Dog or Cat?* was related to the dependent variable *Happiness* at a statistically significant level. Furthermore, analysis indicated that study participants that reported living with a *dog* evidenced a higher level of *Happiness* relative to those living with a *cat*.

Figure 4.119 What multivariate analysis told us for sample study one



Lastly, in terms of magnitude the partial Eta squared effect size estimate indicated that the independent variable *Do You Live with a Dog or Cat?* evidenced a medium to large .09 (Eta squared effect size small, medium, large, = 0.01, 0.06, 0.14, respectively) effect on the dependent variable *Happiness*. Within the context of the full multivariate linear regression model, the strongest predictor of the dependent variable *Happiness* was the covariate predictor variable *Income Level*. This was evident as the covariate

predictor *Income Level* evidenced the strongest standardized beta within the model (β=.39), as well as the largest effect size (PES=.65).

Multivariate Analysis for Sample Study 2

Sample study two employs multivariate analysis to identify the strongest predictor of being *Happy* (Happy=Yes) among the predictor variables (covariate and independent variables) that are significantly related to the dependent variable at the bivariate level ($p<.05$). Within this study, bivariate analysis identified that both covariate predictor variables *Education Level* and *Income Level*, as well as the independent variable *Do You Live with a Dog or a Cat?*, are significantly related to the dependent variable at the bivariate level and will be included in the multivariate analysis.

Figure 5.19 (entered below, as well as on page 269 of TSSODA) presents the **Multivariate Test Key with the variables for sample study two plugged in**. This key describes the appropriate multivariate procedure when examining a dichotomous categorical dependent variable, such as *Happy* (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.

Figure 5.19 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

Before performing the binary logistic regression, it is important to confirm that all necessary test assumptions have been addressed. For example, **Section 5.4.2 Checks of Test Assumptions** (see page 253 of TSSODA text) described the need to assess that no significant **multicollinearity** exists between predictors for this regression model, which was substantiated. A second assumption that must be met is that all predictor variables in the regression model are at the continuous or dichotomous level of measurement. Since the categorical covariate predictor variable *Education Level* has a three category response, the categories within that variable must be dummy-coded (as presented in **Section 5.7.1**) prior to use in the multivariate model. After the process of dummy-

coding has been addressed, the data will be appropriate for use in the binary logistic regression model.

Conducting Binary Logistic Regression

Pages 270-276 describes the process of conducting a Binary Regression model in SPSS. **SPSS Output 5.6** (presented within the subsequent pages, as well as on page 275 of TSSODA) 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 dependent variable *Happy* 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 5.6**, 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 (58.555) 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 85%. 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 dummy-coded variable *EducationCollege* has a probability level of .732, while *EducationGraduateSchool* has a probability level of .000. Remember these results *are in reference* to the reference group, which is *EducationHighSchoolGED*. Therefore, these probability values indicate that the *EducationCollege* group (p=.732) is not significantly more likely to be *Happy* (Yes) relative to the *EducationHighSchoolGED* group. However, the *EducationGraduateSchool* group is significantly (p<.05) more likely to be *Happy* (Yes) relative to the *EducationHighSchoolGED* group as the **Sig.** is **.000** (p<001).

The next line within the **Sig.** column reflects the relationship between the continuous covariate variable *Income Level* and the dependent variable *Happy* (Yes or No) is statistically significant, as the value (**Sig.=.003**) is less than .05 ($p<.05$). The fourth line in this column indicates the relationship between the independent variable *Do You Live with a Dog or a Cat?* and the dependent variable *Happy* (Yes or No) is statistically significant as the value (**Sig.=.002**) is below .05 ($p<.05$).

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**.** Recall, when the beta is preceded by a minus, a negative relationship is indicated, while the absence of a minus sign reflects a positive relationship. For example, the second row within this column (we do not consider the first line as $p>.05$) indicates a beta of 3.17, which indicates a positive relationship where the *EducationGraduateSchool* group is significantly more likely (positive indicates more likely) to be *Happy* (*Happy*=Yes) in reference to the *EducationHighSchoolGED* group. The third line presenting *Income Level*, evidences a positive beta of .04, which indicates as *Income Level* increases the study participant is more likely to be *Happy* (Yes).

Lastly, the fourth line presents the directionality between the independent variable *Do You Live with a Dog or a Cat?* and the dependent variable *Happy* (Yes or No). Observe the beta (B= -2.26) is negative, indicating a negative relationship reflecting as the code for the variable *Do You Live with a Dog or a Cat?* increases from 1 to 2, (1=dog, 2=cat), the code for the dependent variable decreases from 1 to 0, which signals a movement from *Happy* (1=Yes) to not *Happy* (0=No). Thus, there is a significant negative relationship between living with a cat (Cat=2) and being Happy (*Happy*=1).

Part C Relationship Dimension 3: Magnitude

As previously mentioned in logistic regression, the effect size 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.

In sample study two, the odds ratios can be found within the box labeled **Variables in Equation (B)** in the column marked **Exp(B)**. Within this column, on the second line, we see the odd ratio for the dummy-coded variable *EducationGraduateSchool* is 23.85. This value suggests that study participants in the group *EducationGraduateSchool* are 23.85 times more likely to be *Happy* (Yes), in reference to study participants in the *EducationHighSchoolGED* group (the reference group). Thus, a great magnitude is indicated as there is a large effect size between the variable *Education Level* and *Happy* (Yes or No). However, the 95% confidence interval for this odds ratio is extremely wide (95% CI = 4.19-135.67), suggesting the finding is unstable.

Next, on the third line, within the column **Exp(B)**, we see an odds ratio of 1.04 for the covariate variable *Income Level.* 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*. Specifically, when a predictor is categorical (e.g., *Education Level*) in binary logistic regression, it is appropriate to report a greater likelihood of experiencing the event. For example, we could state that those with a *Graduate School* level education are 23.85 times more likely to be *Happy* relative to the reference group.

However, when the predictor variable is continuous, such as *Income Level*, the odds ratio *is not interpretable in the same way, where one might state an 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 *Income level*, we might choose to simply report that a higher *Income Level* is significantly associated with experiencing the event, being *Happy* (Yes), as well as the odds ratio (OR=1.04). However, we might 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 *Do You Live with a Dog or Cat?* and the dependent variable *Happy* (Yes or No), which is .105. This odds ratio is different than the other odds ratios, as the value is below 1.00 (OR=.105). *An odds ratio below 1.0 suggests a negative association with the outcome reflected in the dependent variable*. For example, an odds ratio below 1.0 suggests a negative relationship with experiencing the outcome of being *Happy* (*Happy*=Yes). However, an odds ratio below is not easily

SPSS Output 5.6

Logistic Regression

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	58.555	4	.000
	Block	58.555	4	.000
	Model	58.555	4	.000

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

Classification Table[a]

Observed		Predicted		
		Is the study participant happy?		Percent Correct
		No	Yes	
Is the study participant happy?	Yes	62	7	89.9
	No	8	23	74.2
Overall percent				85.0

Percentage of cases categorized correctly

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

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

Variables in Equation (A)		B	S.E.	Wald	df	Sig.
Step 1	EducationCollege	.33	.95	.12	1	.732
	EducationGraduateSchool	3.17	.88	12.78	1	.000
	Income Level	.04	.01	9.00	1	.003
	Livedogcat (**dog=1, cat=2**)	-2.26	.74	9.32	1	.002
	Constant	2.40	1.19	4.09	1	.043

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

Variables in Equation (B)		Exp (B)	95% Confidence Interval of the Difference	
			Lower	Upper
Step 1	EducationCollege	1.39	.21	9.01
	EducationGraduateSchool	23.85	4.19	135.67
	Income Level	1.04	1.01	3.17
	Livedogcat (**dog=1, cat=2**)	.105	.03	.45
	Constant	11.05		

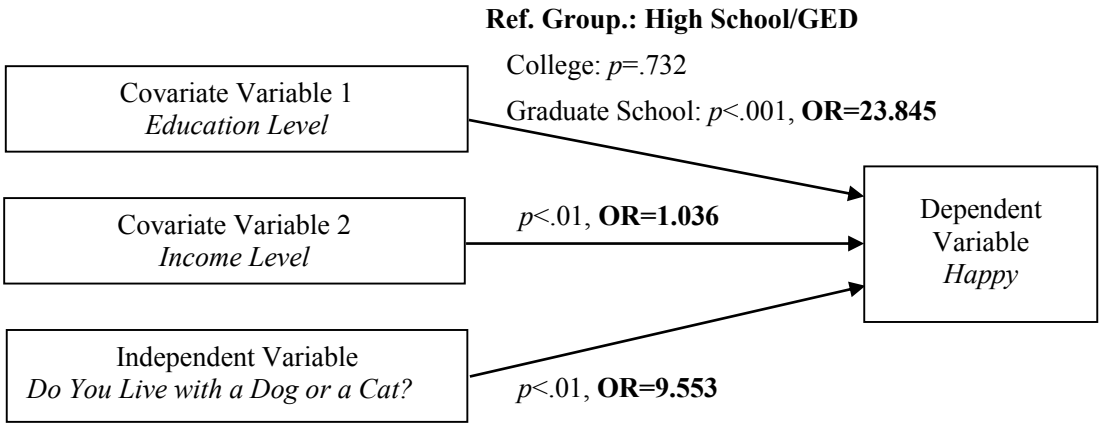
interpreted, so many times these odds ratios are converted to a number greater than 1.0 to better communicate the magnitude of the variable relationship.

Converting a Negative Odds Ratio Manually. To convert an odds ratio below 1.0 manually simply divide 1.0 by the odds ratio. For example, if we divide 1 by the odds ratio .105, the resulting number is 9.5 (1.0/.105=9.5). This suggests that study participants that live with a dog are 9.5 times more likely to be coded as *Happy* (*Happy*=Yes), relative to study participants that live with a cat.

What Did Multivariate Analysis Tell Us?

Figure 5.22 (presented below, as well as on page 277 of TSSODA) describes the findings within the multivariate analysis for sample study two. Before observing the figure, it is important to note that data indicated the overall logistic regression model was statistically significant. Each predictor variable was significantly related to the dependent variable, as indicated within this figure by the probability notation on each of the arrows. Specifically, study participants with a *Graduate School* level education were significantly more likely to be Happy (Yes), relative to the *High School/GED* group, but not the group with a *College* level education. Furthermore, an increase in *Income Level* was significantly associated with being coded as *Happy* (Yes). Lastly, study participants that lived with a dog were significantly more likely to be *Happy* (Yes) relative to those that live with a cat.

Figure 5.22 What multivariate analysis told us for sample study two



In terms of the magnitude of the relationships, we noted large odds ratio effect sizes. For example, study participants with a *Graduate School* level education were described as being 23.85 times more likely to be *Happy* (Yes), relative to the *High School/GED*

group. Additionally, study participants that lived with a dog were almost ten times (9.55) more likely to be *Happy* (Yes) relative to those that live with a cat.

However, as we noted, the 95% confidence intervals were rather wide for these estimates, reflecting that these odds ratios might be rather unstable numbers. The confidence interval for the odds ratio regarding levels of education was especially wide at 4.19-135.67. This wide interval might suggest that although the odds ratio for this group (OR=23.85) is larger than the independent variable (OR=9.55), the relationship reflected between the independent and dependent variable might be more believable.