Week 4:
Univariate Analysis
&
Bivariate Analysis

The Steps of Data Analysis

6 Week Webinar Class

Presented By

William Bannon, PhD

StatsWhisperer

New York, New York

Univariate Analysis

Univariate analysis explores each variable in a dataset separately. Univariate analysis, presented from pages 182-187 in *The 7 Steps of Data Analysis* (TSSODA) textbook, 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 4.88** presented on page 183 of TSSODA textbook, as well as below, 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 4.88 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 4.89** also on page 183 of TSSODA, as well as below, which presents the *Univariate Test Key* with the study variables from our sample study plugged in. Once we have identified the variables used in the sample study as categorical (*Education Level, Do You Live with a Dog or Cat?*) and continuous (*Income Level*, *Happiness*), 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 4.89 Univariate Test Key (Study Variables Plugged In)

Variable	Test to Use (Values to Report)
Categorical	Frequencies (Number & Percentage)
Education Level	Frequencies (Number & Percentage)
Lives w/a Dog/Cat	Frequencies (Number & Percentage)
Continuous	Frequencies (Mean, Std. Deviation, Min/Max)
Income Level	Frequencies (Mean, Std. Deviation, Min/Max)
Happiness	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 *Education Level* (High School/GED, College, and Graduate School).

Continuous Variables

As stated earlier, computing univariate statistics for continuous variables largely involves measures of central tendency and dispersion. Recall in **Figure 4.88** 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.

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 (please see pages 187-203 of TSSODA textbook for a full description of 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 the next step, which is **Multivariate Analysis**.

For example, in sample study one, a goal of bivariate analysis would be to identify which study predictor variables (*Education Level, Income Level, Do You Live with a Dog or Cat?*) are significantly related to the dependent variable *Happiness* 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 *Happiness*. Any predictor variables not significantly related to the dependent variable will be excluded from the multivariate analysis.

Figure 4.91 presents the *Bivariate Test Key* (presented on page 188 of TSSODA textbook, as well as following this paragraph), which describes the appropriate statistical procedure to test the association between two variables, will depend on the structure of each variable. **Figure 4.92** presents the *Bivariate Test Key* with the study variables plugged in (also presented on page 188 of TSSODA textbook, as well as following this paragraph). The first line instructs if Variable 1 is categorical with a two category response, such as *Do You Live with a Dog or a Cat?* (Response: Dog, Cat) and Variable 2 is continuous, such as *Happiness* (Response: 1-5), 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 *Education Level* (Response: High School/GED, College, or

Graduate School) and Variable 2 is continuous, such as *Happiness* (Response: 1-5), 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 *Income Level* (Response: $100,000-$500,000) and Variable 2 is continuous, such as *Happiness* (Response: 1-5), then a **Correlation** would be the appropriate statistical test to examine this association. Sample study one does not require the chi square analysis listed in the fourth line, but this test will be reviewed in detail in sample study two.

Figure 4.91 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 4.92 Bivariate Test Key (Study Variables Plugged In)

Variable 1	Variable 2	Test To Use
Categorical (2 categories)	Continuous	Independent Samples T-Test
Live w/Dog/Cat (Dog, Cat)	Happiness (1-5)	Independent Samples T-Test
Categorical (≥3 categories)	Continuous	One-Way ANOVA
Education (HS/G, C, GS)	Happiness (1-5)	One-Way ANOVA
Continuous	Continuous	Correlation
Income ($100k - $500k)	Happiness (1-5)	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 *Education Level* (Response Categories: High School/GED, College, Graduate School) with the continuous dependent variable *Happiness* (Response Range: 1-5). This procedure will indicate if there is a statistically significant difference in mean *Happiness* scores among study participants with a High School/GED, College, or Graduate School level of education.

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). These assumptions were examined in **Step 3: Checks of Data Integrity**.

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. Pages 188-195 describe how to implement the One-Way ANOVA in SPSS.

Interpreting the Statistical Output: One-Way ANOVA. **SPSS Output 4.26** (presented on page 192 of TSSODA, as well as following Part A of this section) 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 *Happiness* and the covariate variable *Education Level*. 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 **.000**. 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.

SPSS Output 4.26

Oneway

(2nd) Directionality:

Higher education level = Higher *Happiness*

Descriptives
Happiness

	N	Mean	SD	SE	95% CI for the Mean		Min	Max
					Lower	Upper		
High School/GED	32	2.17	.70	.12	1.92	2.42	1.2	4.2
College	29	2.70	.61	.11	2.47	2.93	2.0	4.4
Graduate School	39	3.23	.55	.09	23.05	3.40	2.0	4.4
Total	100	2.73	.76	.08	2.58	2.88	1.2	4.2

(1st) Significance (Model):

Statistically significant (p<.05)

ANOVA
Happiness

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.69	2	9.85	25.96	.000
Within Groups	36.79	97	.38		
Total	56.48	99			

Post Hoc Tests

(1st) Significance (Mean Scores):

Statistically significant (p<.05)

(Mean score differences)

Multiple Comparisons
Dependent Variable: Happiness
Bonferroni

(I) Education Level	(J) Education Level	Mean Difference (I-J)	Std. Error	Sig.	95% CI for the Mean	
					Lower	Upper
High School/GED	College	-.52780*	.16	.004	-.91	-.14
	Graduate School	-1.05689*	.15	.000	-1.41	-.70
College	High School/GED	.52780*	.16	.004	.14	.91
	Graduate School	-.52909*	.15	.002	-.90	-.16
Graduate School	High School/GED	1.05689*	.15	.000	.70	1.41
	College	.52909*	.15	.002	.16	.90

*The mean difference is significant at the 0.05 level.

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 *Happiness* by each response category within the variable *Education Level*. We see that the mean score for *Happiness* for study participants with a High School/GED (*M*=2.17, *SD*=.69) and College (*M*=2.70, *SD*=.61), seem to be much lower relative to that of Graduate School (*M*=3.23, *SD*=.55). 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) Education Level**, presents each category separately. The second column labeled **(J) Education Level**, 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 *Happiness* scores for the study participants with a High School/GED education level, are significantly lower relative to those with a College (-.52780) and Graduate School (-.05689) level of education. In the next row, we see that mean value *Happiness* scores for the study participants with a College education level, are significantly higher relative to those with a High School/GED (.52780) and lower than Graduate School (-.52909) level of education. Lastly, the final row describes mean value *Happiness* scores for study participants with a Graduate School education level, are significantly higher relative to a High School/GED (1.05689) and College (.52909) level of education.

Part C Relationship Dimension 3: Magnitude

Our text explains that 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. Instruction on performing the ANCOVA is presented on pages 192-194 of TSSODA textbook.

SPSS Output 4.27 (presented within our text, as well as following this paragraph) 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 *educ* (*Education Level*) the effect size estimate of the PES is .349. Recall, the effect sizes for the partial Eta squared are

small (>0.01), medium (>0.06), and large (>0.14). Thus, the predictor *Education Level* has a very large effect size (.349) on the dependent variable *Happiness*.

SPSS Output 4.27

Tests of Between-Subjects Effects

Dependent Variable: Happiness

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	16.69	2	9.85	25.96	.000	.349
Intercept	716.44	1	716	1888	.000	.951
Educ	19.69	1	9.85	25.96	.000	.349
Error	36.79	97	.38			
Total	809.96	100				
Corrected Total	56.484	99				

(3rd) Magnitude:

Large effect size (PES = .349)

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 less than .001 ($p<.001$) as the level of probability was .000 (Sig.=.000). In terms of directionality, the mean scores, as well as the post hoc test indicated that a higher *Education Level* (Graduate School: $M=3.23$, $SD=.55$), was associated with a higher level of *Happiness*, relative to a lower Education Level (College: $M=2.70$, $SD=.61$; High School/GED: $M=2.17$, $SD=.70$). Lastly, in terms of magnitude, the *partial Eta squared* effect size (.35) indicated that covariate variable *Education Level* evidenced large effect upon the dependent variable *Happiness*.

Correlation

The Bivariate Test Key indicates a correlation is the appropriate statistical procedure examining the relationship between a continuous variable with another continuous variable (please see pages 195-197 of our textbook for the correlation procedure in detail). Therefore, we will use this procedure to examine the relationship between the continuous covariate variable *Income Level* (Response Range: $100k-$500k) with the continuous dependent variable *Happiness* (Response Range: 1-5). This procedure will indicate if *Happiness* scores either increase or decrease at a statistically significant level as *Income Levels* 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). These assumptions were examined in **Step 3: Checks of Data Integrity**. Pages 198-199 presents the steps to conduct the Pearson's correlation in SPSS.

Interpreting the Statistical Output: Pearson's Correlation. **SPSS Output 4.28** (presented on page 197 of TTSODA, as well as on the following page) 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 *Happiness* and the covariate variable *Income Level*. 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 cell at the end of the first row that begins with the variable name *Income Level*. Within this row, first refer to the second line labeled **Sig. (2-tailed)**. This line presents the **statistical significance** (probability level) for the procedure, which is .000 (in the far right hand column), indicating that the test 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 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, the relationship between the variables is negative. However, there is not a minus sign, which indicates a positive relationship whereas scores reflecting the variable *Income Level* increases, scores within the variable *Happiness* increase.

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. The correlation coefficient can also be squared to produce an **effect size** (reflecting the variance in one variable explained by the other variable). If

the coefficient .884 is squared the resulting value is .7815, indicating that over 78% (78.15%) of the variance in the dependent variable *Happiness* is explained by the covariate variable *Income Level*. This indicates a large effect size as guidelines for the social sciences suggest a small, medium, and large effect size for the r^2 to be 0.10, 0.30, and 0.50 (r^2 to be 0.01, 0.09, 0.25), respectively (Cohen, 1992).

SPSS Output 4.28

Correlations

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

(3rd) Magnitude:
Large effect size (r^2 = .78)

(1st) Significance:
Statistically significant

Correlations

		Income Level	Happiness
Income Level	Pearson Correlation	1	.884**
	Sig. (2-tailed)		.000
	N	100	100
Happ-iness	Pearson Correlation	.884**	1
	Sig. (2-tailed)	.000	
	N	100	100

**Significant at the 0.01 level (2-tailed).

Summary. Overall, the Pearson's correlation 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 less than .001 (p<.001) as the level of probability was .000 (Sig.=.000). In terms of directionality, the correlation coefficient was positive, which indicates a positive relationship where *Happiness* scores increase as *Income Level* increases. Lastly, in terms of magnitude, after squaring the correlation coefficient (r^2 = .78 effect size) we saw that the covariate variable *Income Level* evidenced a large effect upon the dependent variable *Happiness*.

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 (please see pages 197-202 of TSSODA for a detailed presentation of this procedure). Therefore, we will use this procedure to examine the relationship between the categorical independent variable *Do You Live with a Dog or Cat?* with the continuous dependent variable *Happiness*. This procedure will indicate if there is a statistically significant difference between mean *Happiness* scores (Response Range: 1-5), among study participants that live with a dog or cat.

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). These assumptions were examined in **Step 3: Checks of Data Integrity**.

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. Pages 198-199 of TSSODA text presents the steps to conduct the independent Samples t-test in SPSS.

Interpreting the Statistical Output: Independent-Samples T-Test. **SPSS Output 4.29** (presented on page 200 of TSSODA, as well as on the following page) 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 *Happiness* and the independent variable *Do You Live with a Dog or a Cat?* 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 (*Happiness*) are significantly unequal among the two categorical response groups (lives with dog, lives with cat). 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.=.872, 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 *Happiness* scores for the two categorical groups included in the t-test procedure (study participants that live with a *dog* or *cat*). Within the column labeled *Do You Live with a Dog or Cat?* the first row labeled *Dog* reflects there are 47 study participants that live with a dog (*n*=47) with a mean *Happiness* score of 3.02 (*SD*=.79). The second row labeled *Cat* reflects there are 53 study participants that live with a cat (*n*=53) with a mean *Happiness* score of 2.48 (*SD*=.62). Thus, the independent-samples t-test indicates that higher ***Happiness*** **scores are associated with being in the group living with a dog relative to a cat**.

SPSS Output 4.29

(2nd) Directionality: Living with a Dog = Higher *Happiness*

(3rd) Magnitude: Large Effect size (Cohen's *d*= 0.76)

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

T-Test

Group Statistics

	Do You Live with a Dog or a Cat?	N	Mean	Standard Deviation	Std. Error Mean
Happiness	Dog	47	3.0213	.79371	.11577
	Cat	53	2.4792	.62307	.08559

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
Happiness	Equal Variances assumed	.03	.87	3.8	98	.000	1.5	.71	2.2
	Equal Variances not assumed			3.8	97.8	.000	1.5	.71	2.3

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 (the formula is presented in **Figure 4.101**). Specifically, to compute Cohen's *d*:

1) **Compute the denominator (the *pooled standard deviation* presented in Figure 4.102**):
    a. Square the **standard deviation** for each group:
    SD of Lives with a Dog: $.79371^2 = .6299755641$
    SD of Lives with a Cat: $.62307^2 = .3882162249$
    b. Sum those values: .6299755641 + .3882162249 = 1.018191789
    c. Divide that summed number by 2: 1.018191789/2 = .5090958945
    d. Take the square root of that number: .71350956160376715

2) **Compute the numerator**: Subtract the mean score of group 1 (3.0213) from the mean of group 2 (2.4792), which is = 0.5421.

3) **Divide the numerator by the denominator**: 0.5421/.71350956160376715 = 0.7598.

Figure 4.101 Formula to compute Cohen's *d*

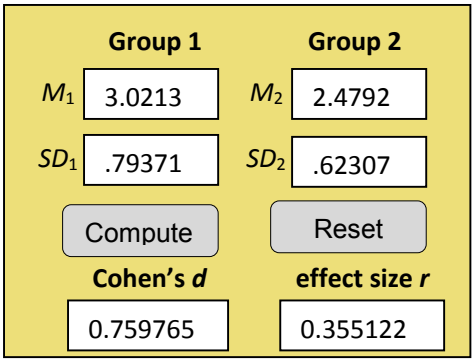
$$d = \frac{Mgroup1 - Mgroup2}{SDpooled}$$

Figure 4.102 The *pooled standard deviation* to compute the Cohen's *d*

$$SDpooled = \sqrt{(SD^2group1 + SD^2group2)/2}$$

An easier 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 4.103** 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 *Do You Live with a Dog or a Cat?* has a large effect on the dependent variable *Happiness*.

Figure 4.103 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 *Happiness* was associated with living with a dog ($M=3.02$, $SD=.79$) relative to a cat ($M=2.48$, $SD=.62$). Lastly, in terms of magnitude, the Cohen's *d* reflected that the independent variable *Do You Live with a Dog or a Cat?* evidenced a large effect (Cohen's *d* effect size=0.7598) upon the dependent variable *Happiness*.

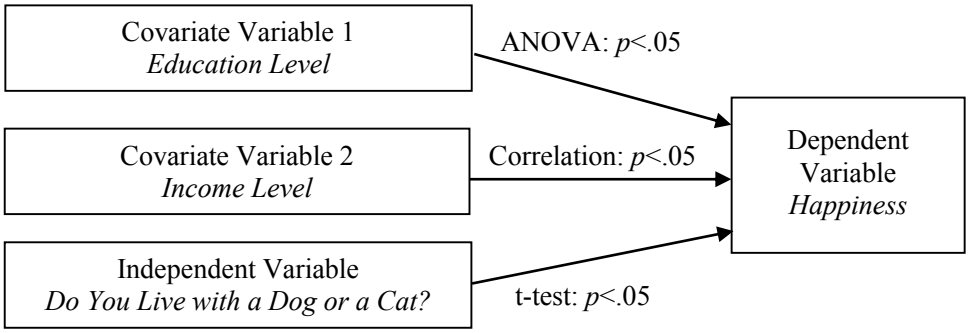
What Did Bivariate Analysis Tell Us?

Within the context of our sample studies, 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 4.104 (presented on page 203 of TSSODA textbook, as well as on the following page) reflects the results on the three bivariate tests used in sample study one examining which predictor variables have a statistically significant association with the continuous dependent variable *Happiness* (1-5). Notice that we have entered an arrow connecting covariate variable 1: *Education Level* 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 between

covariate variable 2: *Income Level* and the independent variable: *Do You Live with a Dog or a Cat?* with the dependent variable.

Figure 4.104 What Bivariate Analysis Tells Us for Sample Study One?



Subsequently, the findings produced through bivariate analysis tells us that all three 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 *Happiness* (1-5). That dimension will be revealed in multivariate analysis.