

Chapter 14

Making Use of Associations Tests

Learning Objectives

- To learn what is meant by an “association” between two variables
- To examine various relationships that may be construed as associations
- To understand where and how cross-tabulations with Chi-square analysis are applied
- To become familiar with the use and interpretation of correlations
- To learn how to obtain and interpret cross-tabulations, Chi-square findings, and correlations with SPSS

"WHERE WE ARE"

- 1 Establish the need for marketing research.
- 2 Define the problem.
- 3 Establish research objectives.
- 4 Determine research design.
- 5 Identify information types and sources.
- 6 Determine methods of accessing data.
- 7 Design data-collection forms.
- 8 Determine the sample plan and size.
- 9 Collect data.
- 10 Analyze data.
- 11 Prepare and present the final research report.

Significant Associations can help Managers make Better Decisions

- **Associative analyses:** determine where stable relationships exist between two variables.
- **Associative analyses** procedures are useful because they determine if there is a consistent and systematic relationship between the presence(label) or amount (level) of one variable and the presence (lable) or amount (level) of another variable.

Types of Relationships Between Two Variables

- **Relationship**: a consistent, systematic linkage between the levels or labels for two variables
 - “**Levels**” refers to the characteristics of description for interval or ratio scales.
 - “**Labels**” refers to the characteristics of description for nominal or ordinal scales.

There are four basic types of Relationships Between Two Variables

1-Nonmonotonic.

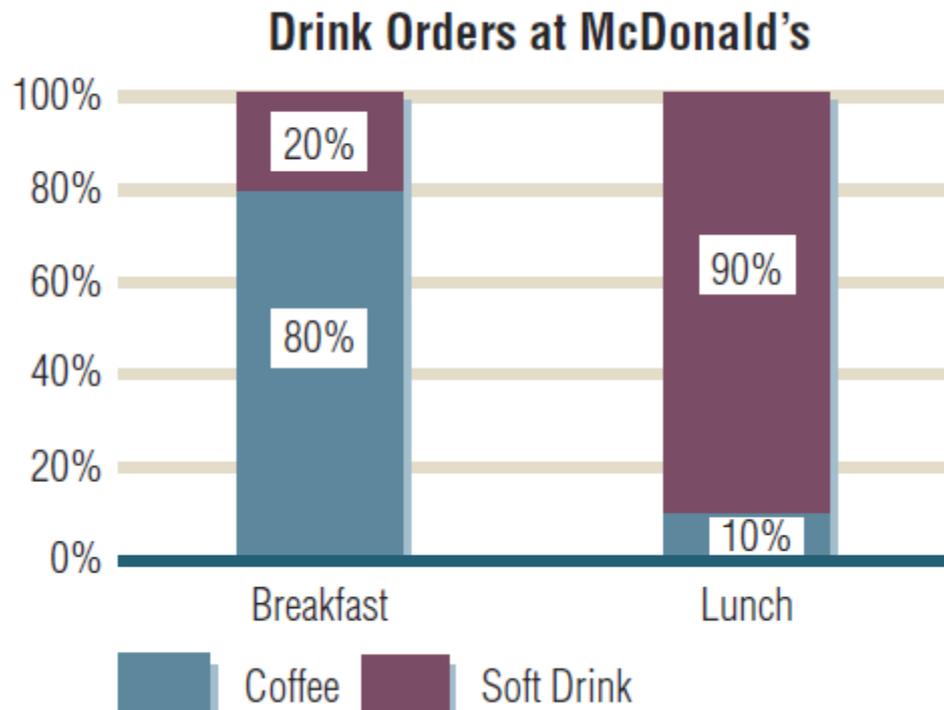
2- Monotonic.

3- Linear.

4- Curvilinear.

1-Nonmonotonic relationship: two variables are associated, but only in a very general sense. The presence (or absence) of one variable is associated with the presence (or absence) of another.

Example Relationship



2-Monotonic relationship: the general direction of a relationship between two variables is known.

- **There are two types of monotonic relationship;**
 - a) Increasing relationship;** are those in which one variable increases as the other variable increases.
 - b) Decreasing relationship;** are those in which one variable increases as the other variable decreases.

- 3- Linear relationship:** “straight-linear association” between two variables.
- Linear relationship are quite precise.

Formula for a straight line

$$y = a + bx$$

Where:

y = the dependent variable being estimated or predicted

a = the intercept

b = the slope

x = the independent variable used to predict the dependent variable

- 4- Curvilinear relationship:** are those in which one variable is associated with another variable, but the relationship is described by a curve rather than a straight line.
- The relationship may be an S-shape, a J-shape, or some other curved-shape pattern.

Characterizing Relationships Between Variables

- Depending on its type, a relationship can usually be characterized in three ways by ;
 - 1-Presence:** whether any systematic (statistical) relationship exists between two variables
 - 2-Direction or pattern:** **Direction** means that you know if the relationship is positive or negative , While **pattern** means you know the general nature of the relationships.
 - 3-Strength of association:** how strong the relationship is : strong? Moderate? Weak?

TABLE 14.1 Step-by-Step Procedure for Analyzing Relationships

Step	Description
1. Choose variables to analyze.	Identify which variables you think might be related.
2. Determine the scaling assumptions of the chosen variables.	For purposes of this chapter, both must be either scale (interval or ratio) or categorical (nominal) variables.
3. Use the correct relationship analysis.	For two nominal variables, use cross-tabulation; for two scale variables, use correlation.
4. Determine if the relationship is present.	If the analysis shows relationship is statistically significant, it is present.
5. If present, determine the direction of the relationship.	A linear (scale variables) relationship will be either increasing or decreasing; a nonmonotonic relationship (nominal scales) will require looking for a pattern.
6. If present, assess the strength of the relationship.	With correlation, the size of the coefficient denotes strength; with cross-tabulation, the pattern is subjectively assessed.

Cross-Tabulations

- Cross-tabulation and the Associated Chi-square value we are about to explain are used to assess if a nonmonotonic relationship exists between two nominally scaled variables.
- **Cross-tabulation:** rows and columns defined by the categories classifying each variable; used for nonmonotonic relationships
- **Cross-tabulation cell:** the intersection of a row and a column

Types of Frequencies and percentages in a Cross-Tabulations Table

- **Frequencies** are the raw numbers in the cell.
- **Three types of numbers in each cell**
 - 1-Row percentage
 - 2-Column percentage
 - 3-Row percentage

1-Row percentages table are cell frequencies divided by the grand total.

2-Row percentages table are the row cell frequencies divided by its row total.

3-Column percentages table are the column cell frequencies divided by its column total.

***Column cell percent = Cell frequency / Total of cell frequencies in that column**

***Row cell percent = cell frequency / Total of cell frequencies in that row**

TABLE 14.2A Cross-Tabulation Frequencies Table for a Michelob Light Survey

		Type of Buyer			Totals
		Buyer	Nonbuyer		
Occupational Status	White Collar	152	+	8	= 160
	Blue Collar	14	+	26	= 40
Totals		166	+	34	= 200

Buyer/White Collar Cell Frequency	→	152			
Row Totals	←				160
Column Totals	→	166			
Grand Total	→			34	200

TABLE 14.2B Cross-Tabulation Percentages Tables for a Michelob Light Survey

Raw Percentages Table

		Buyer		Nonbuyer		Totals
Occupational Status	White Collar	76% (152/200)	+	4% (8/200)	=	80% (160/200)
	Blue Collar	7% (14/200)	+	13% (26/200)	=	20% (40/200)
Totals		83% (166/200)	+	17% (34/200)	=	100% (200/200)

Column Percentages Table

		Buyer		Nonbuyer		Totals
Occupational Status	White Collar	92% (152/166)	+	24% (8/34)	=	80% (160/200)
	Blue Collar	8% (14/166)	+	76% (26/34)	=	20% (40/200)
Totals		100% (166/166)		100% (34)		100% (200/200)

Row Percentages Table

		Buyer		Nonbuyer		Totals
Occupational Status	White Collar	95% (152/160)	+	5% (8/160)	=	100% (160/160)
	Blue Collar	35% (14/40)	+	65% (26/40)	=	100% (40/40)
Totals		83% (166/200)	+	17% (34/200)	=	100% (200/200)

Chi-Square Analysis

- **Chi-square analysis:** the examination of frequencies for two nominal-scaled variables in a cross-tabulation table to determine whether the variables have a significant relationship

- Assesses non-monotonic association in a cross-tabulation table based upon differences between observed and expected frequencies.
 - The null hypothesis is that the two variables are not related.
 - Observed frequencies are the actual cell counts in the cross-tabulation table.
 - Observed frequencies are compared to expected frequencies.

Observed and Expected Frequencies

- **Observed Frequencies** , which are the actual cell count in the cross-tabulation table.
- **Expected frequencies** which are defined as the theoretical frequencies in each cell that are derived from this hypothesis of no association between the two variables.

$$\text{Expected cell frequency} = \frac{\text{Cell column total} \times \text{Cell row total}}{\text{Grand total}}$$

The Computed χ^2 Value

- Computed Chi-square values:

$$\chi^2 = \sum_{i=1}^n \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

where

Observed_{*i*} = observed frequency in cell *i*

Expected_{*i*} = expected frequency in cell *i*

n = number of cells

- The computed Chi-square value compares observed to expected frequencies.
- **The Chi-square statistic summarizes** is the degree to which the observed frequencies depart from the expected frequencies is express in single number.

The Chi-Square Distribution

Formula for Chi-Square
degrees of freedom

$$\text{Degrees of freedom} = (r - 1)(c - 1)$$

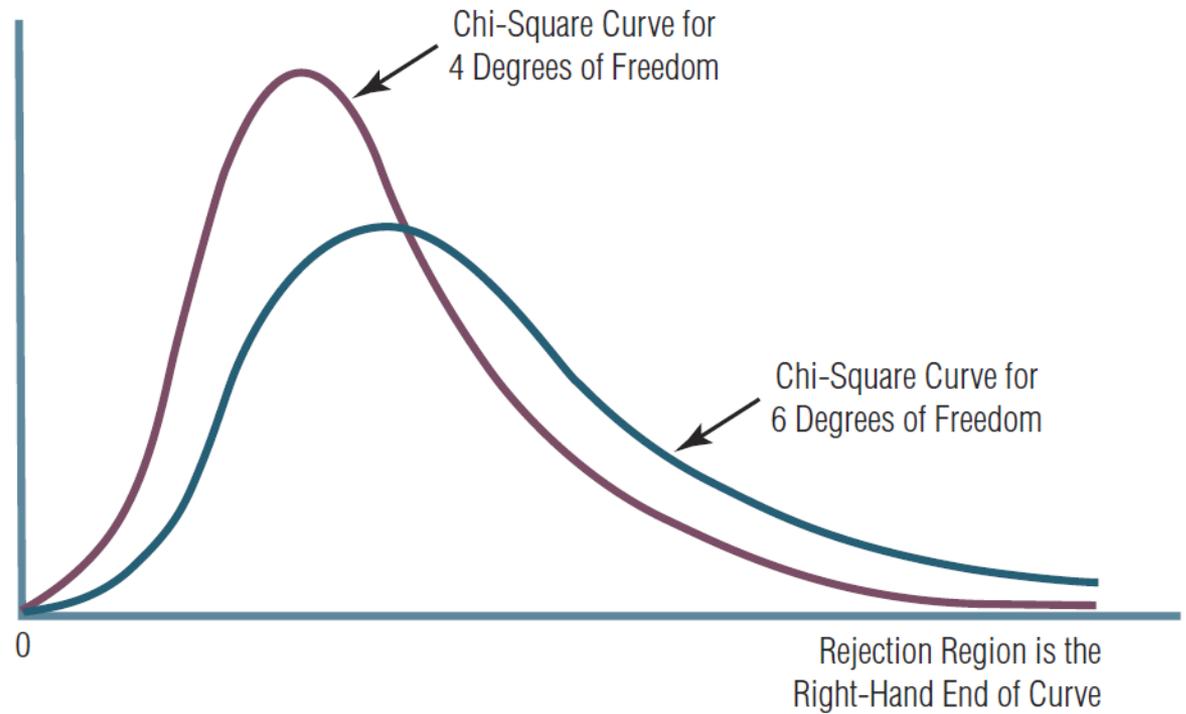
Where:

r = the number of rows

c = the number of columns

- **The Chi-square distribution** is skewed to the right, and the rejection region is always at the right-hand tail of the distribution.
- The shape of the distribution is dependent on degrees of freedom.
- The computed Chi-Square value is compared to a table value to determine statistical significance.
- Computer statistical programs look up table Chi-Square values and print out the probability of support for the null hypothesis.

FIGURE 14.2 The Chi-Square Curve's Shape Depends on its Degrees of Freedom



How to Interpret a Chi-square Result

- Chi-Square analysis yields the amount of support for the null hypothesis if the researcher repeated the study many times with independent sample.
- SPSS and other statistical analysis programs compare calculated to table values and show the probability for support of the null hypothesis.
- A significant Chi-square means the researcher should look at the cross-tabulation row and column percentages to “see” the association pattern.

- With SPSS, Chi-Square is an option under the “crosstabs” analysis routine .
- With Chi-Square analysis , interpret the SPSS significance level as the amount of support for no association between the two variables being analyzed.

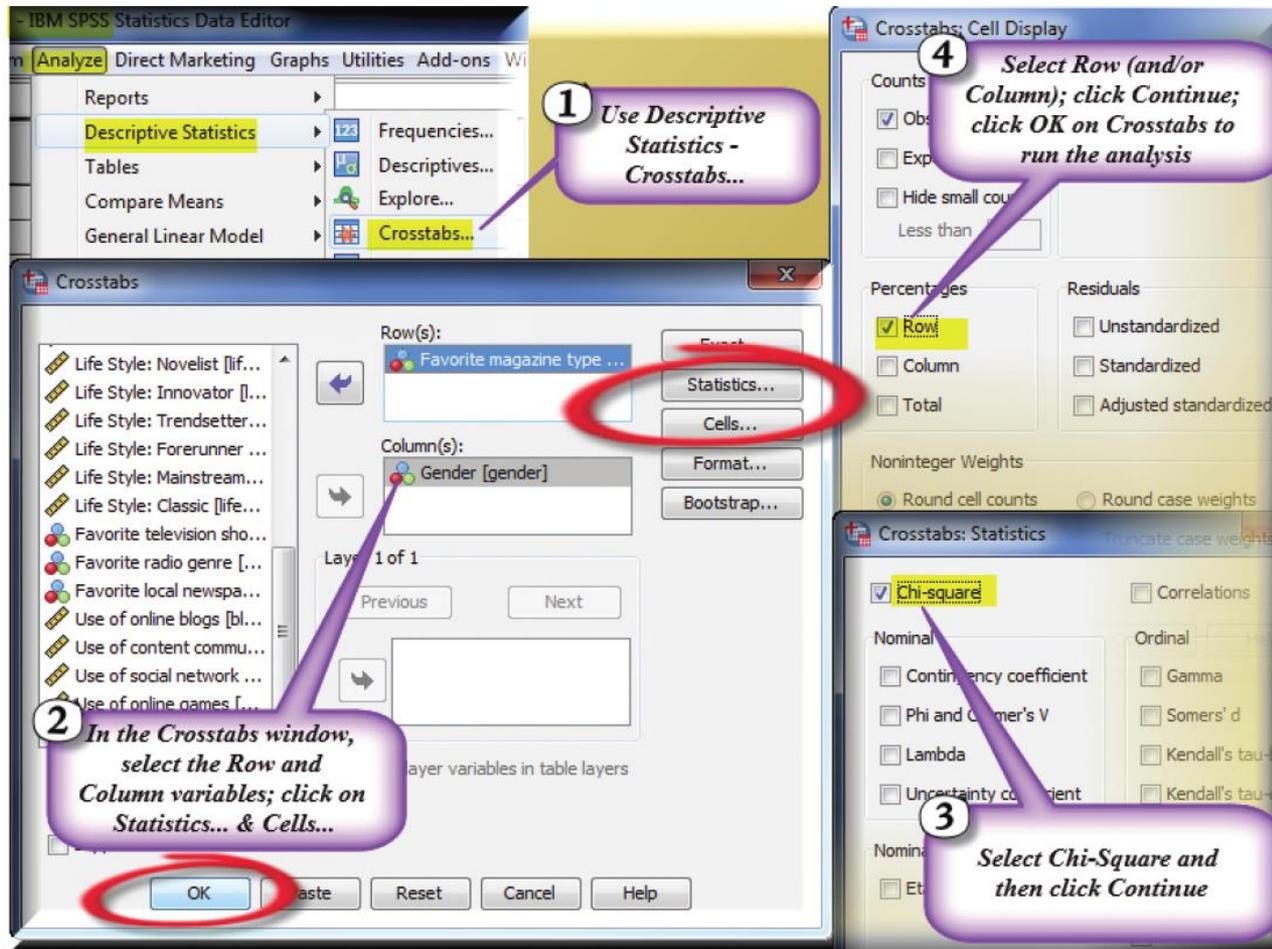
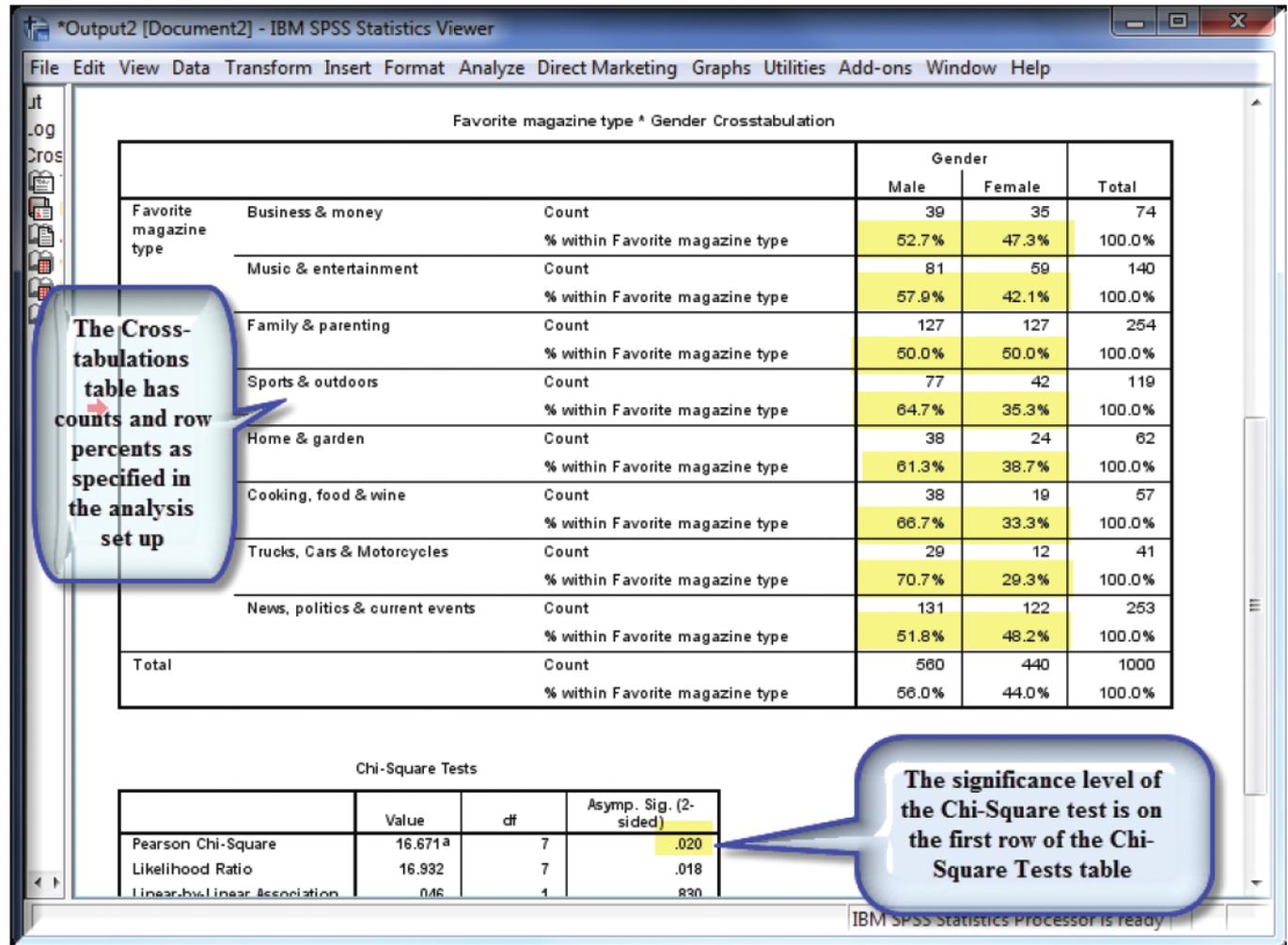


FIGURE 14.3 SPSS Clickstream to Create Cross-Tabulations with Chi-Square Analysis

Source: Reprint courtesy of International Business Machines Corporation, ©SPSS, Inc., an IBM Company.

FIGURE 14.4 SPSS Output for Cross-Tabulations with Chi-Square Analysis



Source: Reprint courtesy of International Business Machines Corporation, ©SPSS, Inc., an IBM Company.

- If the P value is \leq to 0.05, this means there is little support for the null hypothesis (no association).
- Therefore, we have a significant association and have the PRESENCE of a systematic relationship between the two variables.

Reporting Cross-Tabulation Findings to Clients

- Whenever the researcher finds a statistically significant cross-tabulation relationship, he moves to the presentation phase.
- Researcher often turn to graphical presentation as pictures will show the relationships adequately.
- Bar charts can be used to “see” a nonmonotonic relationship.

- **Whether to report row or column percents depends on research objectives.**

Column Percents Table

Size of Sandwich Ordered	Males	Females
<i>Jumbo Size</i>	50%	5%
<i>Large Size</i>	40%	20%
<i>Regular Size</i>	10%	75%
<i>Total</i>	100%	100%

Rows Percents Table

Size of Sandwich Ordered	Males	Females	Total
<i>Jumbo Size</i>	90%	10%	100%
<i>Large Size</i>	67%	33%	100%
<i>Regular Size</i>	13%	87%	100%

Correlation Coefficients and Covariation

- The **correlation coefficient**: an index number, constrained to fall between the range of -1.0 and $+1.0$
- The correlation coefficient communicates both the strength and the direction of the linear relationship between two metric variables.
- To use a correlation, you must first determine that it is statistically significant from zero.
- **Covariation**: the amount of change in one variable systematically associated with a change in another variable

The Correlation Sign the Direct of the Relationship

- A correlation indicates the strength of association between two variables by the absolute size of the correlation coefficient.
- The direction of the association is communicated by the sign (+, -) of the correlation coefficient.

Graphing Covariation using Scatter Diagrams

-Covariation can be examined with use of a scatter diagram.

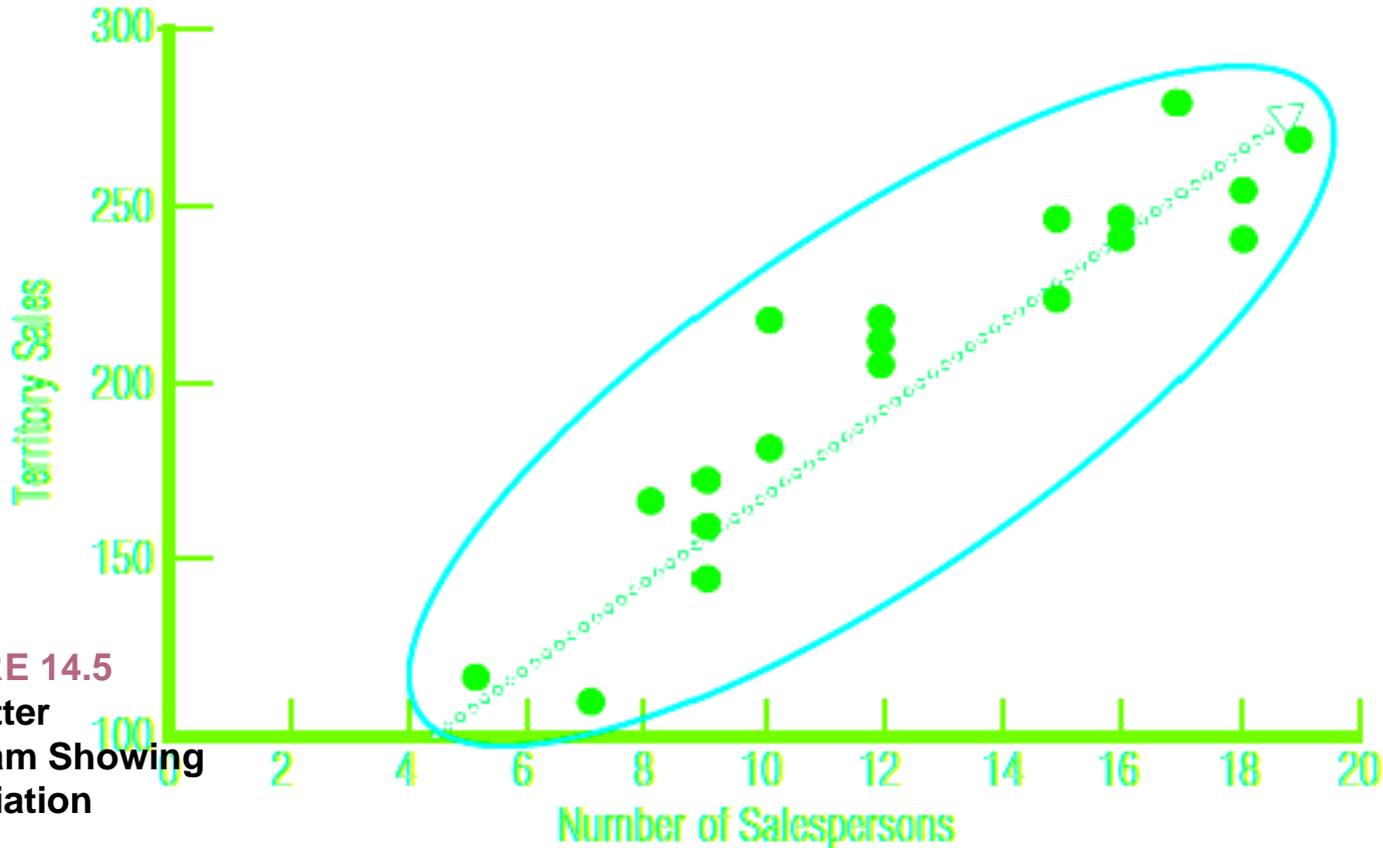
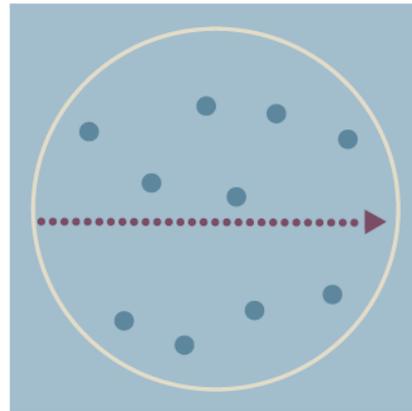


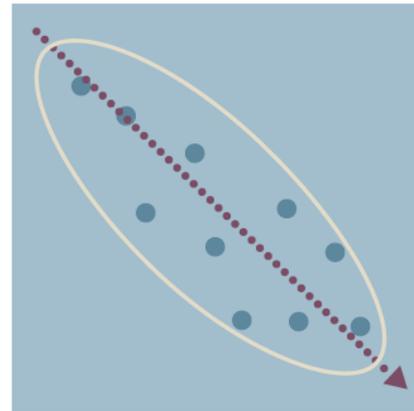
FIGURE 14.5
A Scatter
Diagram Showing
Covariation

Graphing Covariation Using Scatter Diagrams

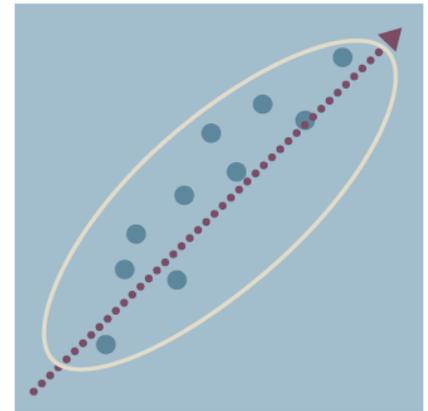
FIGURE 14.6 Scatter Diagrams Illustrating Various Relationships



(a) No Association



(b) Negative Association



(c) Positive Association

- A correlation coefficient's size indicates the strength of association between two variables.
- The sign (+ or -) indicates the direction of the association.

TABLE 14.3 Rules of Thumb about Correlation Coefficient Size*

Coefficient Range	Strength of Association*
+0.81 to +1.00; -0.81 to -1.00	Strong
+0.61 to +0.80; -0.61 to -0.80	Moderate
+0.41 to +0.60; -0.41 to -0.60	Weak
+0.21 to +0.40; -0.21 to -0.40	Very weak
+0.20 to -0.20	None

*Assuming the correlation coefficient is statistically significant

Pearson Correlation Coefficient (r)

- **Pearson Correlation Coefficient (r):**
measures the degree of linear association
between the two variables

where

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{ns_x s_y}$$

x_i = each x value

\bar{x} = mean of the x values

y_i = each y value

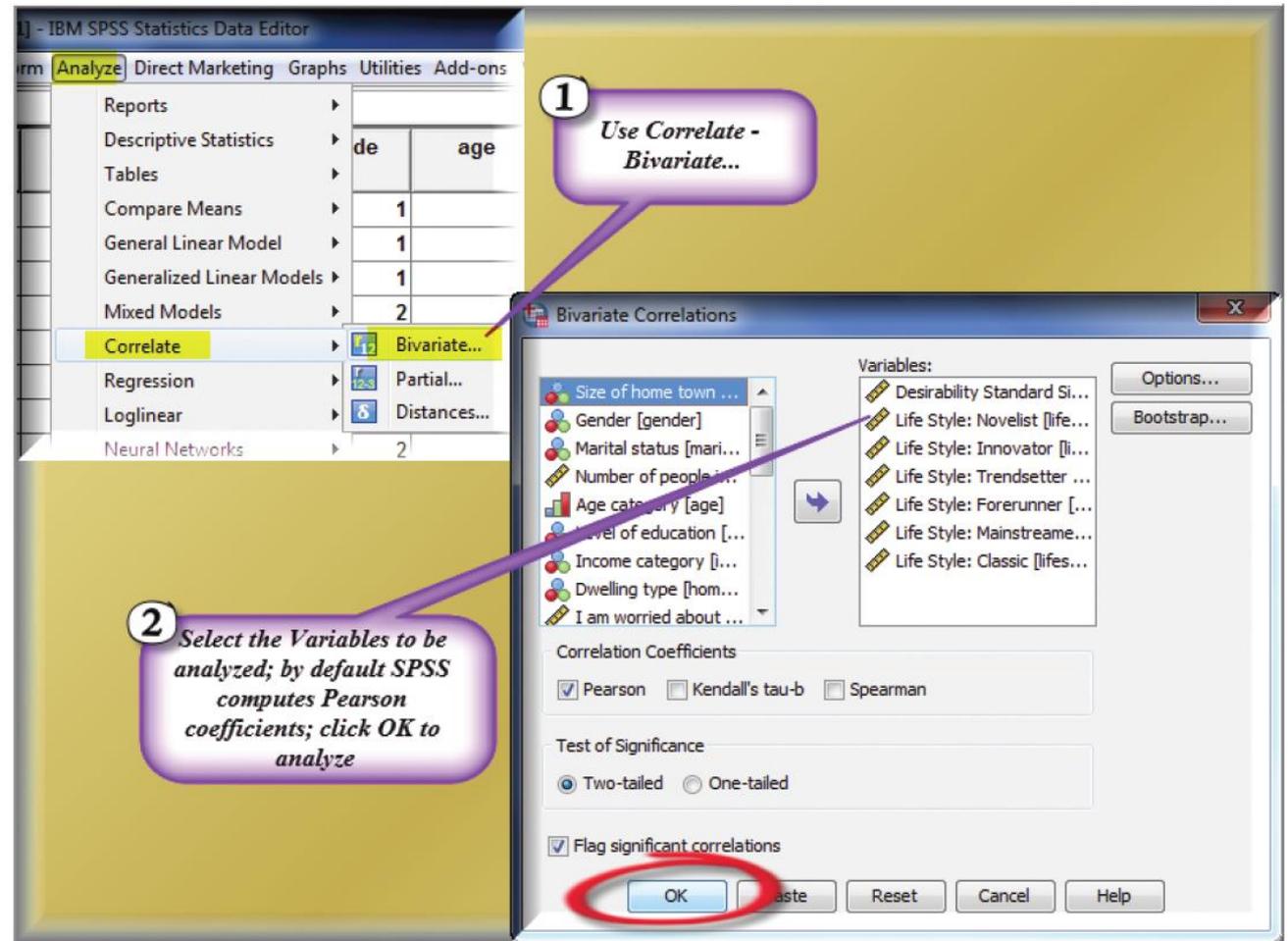
\bar{y} = mean of the y values

n = number of paired cases

s_x, s_y = standard deviations of x and y, respectively

- A positive correlation signals an increasing linear relationship whereas a negative correlation signals a decreasing one.

FIGURE 14.7 SPSS
Clickstream to Obtain
Correlations



Source: Reprint courtesy of International Business Machines Corporation, ©SPSS, Inc., an IBM Company.

FIGURE 14.8
SPSS Output for
Correlations

The Correlations table has correlations, significance levels, and sample size for each pair

The table is symmetric with 1s on the diagonal meaning you use only one-half of it.

		Desirability Standard Size Gasoline	Life Style: Novelist	Life Style: Innovator	Life Style: Trendsetter	Life Style: Forerunner	Life Style: Mainstreamer
Desirability Standard Size Gasoline	Pearson Correlation	1	.090	-.116			
	Sig. (2-tailed)		.004	.000			
	N	1000	1000	1000			
Life Style: Novelist	Pearson Correlation	.090	1	-.061			
	Sig. (2-tailed)	.004	.056				
	N	1000	1000	1000	1000	1000	1000
Life Style: Innovator	Pearson Correlation	-.116	-.061	1	-.165	.003	-.023
	Sig. (2-tailed)	.000	.056	.000	.000	.917	.467
	N	1000	1000	1000	1000	1000	1000
Life Style: Trendsetter	Pearson Correlation	.022	.105	-.165	1	-.010	-.009
	Sig. (2-tailed)	.491	.001	.000	.743	.772	.772
	N	1000	1000	1000	1000	1000	1000
Life Style: Forerunner	Pearson Correlation	.217	.111	.003	-.010	1	.092
	Sig. (2-tailed)	.000	.000	.917	.743	.004	.004
	N	1000	1000	1000	1000	1000	1000
Life Style: Mainstreamer	Pearson Correlation	-.008	.020	-.023	-.009	.092	1
	Sig. (2-tailed)	.795	.536	.467	.772	.004	.004
	N	1000	1000	1000	1000	1000	1000
Life Style: Classic	Pearson Correlation	.634	.070	-.070	.068	.097	-.043
	Sig. (2-tailed)	.000	.026	.027	.032	.002	.177
	N	1000	1000	1000	1000	1000	1000

IBM SPSS Statistics Processor is ready

Source: Reprint courtesy of International Business Machines Corporation, ©SPSS, Inc., an IBM Company.

- Whenever you instruct SPSS to compute correlations, its output is a symmetric correlation matrix composed of rows and columns that pertain to each of the variables.
- Each cell in the matrix contains three items;
 - 1- The correlation coefficient.
 - 2- The significance level.
 - 3- The sample size.
- *With correlation analysis, each correlation has unique significance level.

Special Considerations in Linear Correlation Procedures

- Perhaps because the word **Correlation** is used in everyday language, statistical correlations are sometimes misunderstood by clients.
- **Four caution to keep in mind when working with correlations:**
 - 1-Assumes both variables share interval scaling at minimum.
 - 2-Only takes into account the relationship between two variables, not interaction with other variables.
 - 3-Does not demonstrate cause and effect.
 - 4-Will not detect nonlinear relationships between variables.

Reporting Correlation Findings to Clients

- The researcher must test to determine that a significant correlation has been found before reporting it.
- To our knowledge, there is no marketing research industry standard on how to report statically significant correlations to clients.

Reporting Example

Variables Correlated with Subshop Patronage

Variable	Correlation
<i>Variables positively correlated with patronage:*</i>	
<i>I tend to use the same sandwich shop.**</i>	.76
<i>I worry about calories.**</i>	.65
<i>Age</i>	.55
<i>Number of years with present company</i>	.40
<i>Variables negatively correlated with patronage:*</i>	
<i>I “do” lunch at the place closest to my work.**</i>	-.71
<i>Years of education</i>	-.51

*Subshop Patronage (number of times used in past 2 months)

** Based on a 7-point scale where 1 = strongly disagree and 7 = strongly agree