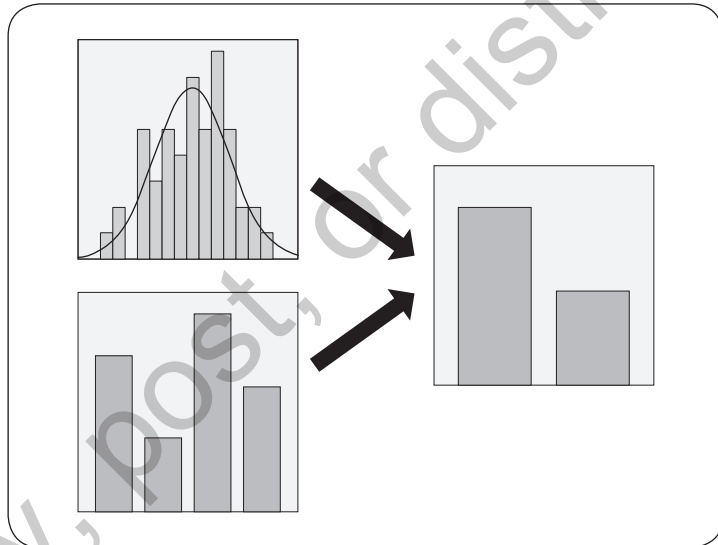# C H A P T E R   1 3

Logistic Regression

Logistic Regression predicts the likelihood of an outcome occurring (or not).

The right choices over time greatly improve your odds of a long and healthy life.

—Tom Rath

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

- Determine when it is appropriate to run a logistic regression analysis
- Verify that the data meet the criteria for logistic regression processing: sample size, normality, and multicollinearity

(Continued)

(Continued)

- Order a logistic regression test
- Comprehend the R^2 statistic
- Label and derive results from the *Variables in the Equation* table
- Selectively process findings to respond to a variety of research questions
- Understand the rationale for recoding categorical variables
- Resolve the hypotheses
- Document the results in plain English
- Comprehend the fundamental principle of multiple regression (R^2)

VIDEO

The tutorial video for this chapter is **Ch 13 – Logistic Regression.mp4.** This video provides an overview of the logistic regression statistic, followed by the SPSS procedures for processing the pretest checklist, ordering the statistical run, and interpreting the results of this test using the data set: **Ch 13 – Example 01 – Logistic Regression.sav.**

OVERVIEW—LOGISTIC REGRESSION

In health science, there are interventions, experiments, and general happenstances that produce dichotomous results, wherein one of two possible outcomes occurs. For example, resuscitation could be thought of as having one of two possible outcomes—the patient either does or does not survive. Other examples that could be considered as having dichotomous outcomes involve did/did not have an adverse drug reaction, did/did not develop a nosocomial infection, did/did not adhere to prescribed medication regime, did/did not have surgical complications, and does/ does not have fever.

In addition to including the outcome variable (e.g., still smoking/quit smoking) in a logistic regression model, it also includes (predictor) variables (e.g., age, income, baseline daily smoking, gender, race); these are variables that are reasonably thought to be associated with the outcome variable. The logistic regression processor assesses the relationships among the variables to provide a model that describes the (predictive) factors associated with the observed outcome.

While the logistic regression model insists on a dichotomous (two-category) outcome variable, you may have surmised from this example that this statistic is liberal in terms of the types of predictor variables that can be included. Logistic regression accommodates continuous predictor variables (e.g., age, income, baseline daily smoking), categorical predictor variables (e.g., gender, race), or any combinations(s) thereof.

The findings from a logistic regression model can provide insights as to the outcome of a current investigation, or in some cases, the findings may serve as a viable predictive model, anticipating the outcome of a future similar circumstance.

Example

A nurse has conducted a smoking cessation workshop for a wide variety of patients who wish to quit smoking. At the conclusion of the series, instead of simply calculating the percentage of the participants who quit smoking, logistic regression is used to better comprehend the characteristics of those who succeeded.

Research Question

What influences do (predictive) variables such as age, income, baseline mean number of cigarettes smoked daily, gender, and race have when it comes to quitting (or not quitting) smoking?

Groups

In this example, all of the members are included in a single group—everyone receives the same smoking cessation intervention.

Procedure

As a public service, the Acme Health Center advertises and offers a free 90-day smoking cessation program, consisting of nurse-facilitated psychoeducational meetings, peer support from those who have been smoke free for more than 1 year, and multimedia resources designed to promote smoking cessation.

At the conclusion of the intervention, each participant is requested to respond to a self-administered anonymous *Smoking Cessation Survey* card (Figure 13.1).

Hypotheses

Considering that this is the first run of this intervention, we have no plausible basis for presuming that any of the predictors will produce statistically significant findings (e.g., females will quit more frequently than males, those with higher income will have a better chance at quitting than those with lower income, etc.). Naturally, such hypotheses could be drafted; however, for this initial run, we will take a less specific exploratory approach:

H_0: Age, income, baseline smoking, gender, and race do not influence one's success in a smoking cessation intervention.

H_1: Age, income, baseline smoking, gender, or race influence one's success in a smoking cessation intervention.

Figure 13.1	*Smoking Cessation Survey* card, anonymously completed by each participant at the conclusion of the intervention.

Smoking Cessation Survey

1. What is your age? _____

2. What is your annual (gross) income? _____

3. Prior to this intervention, how many cigarettes did you smoke in an average day? _____

4. What is your gender?

   ☐ Female   ☐ Male

5. What is your race?

   ☐ African American   ☐ Asian   ☐ Caucasian   ☐ Latino   ☐ Other

6. What is your current smoking status?

   ☐ Still smoking   ☐ Quit smoking

   Please drop this card into the survey box.

   Thank you for your participation.

Data Set

Use the following data set: **Ch 13 – Example 01 – Logistic Regression.sav.**

Codebook

Variable:	Gender
Definition:	Predictor: Gender
Type:	Categorical
	0 = Female [← *REFERENCE*]
	1 = Male
Variable:	Race
Definition:	Predictor: Race
Type:	Categorical
	0 = African American [← *REFERENCE*]
	1 = Asian
	2 = Caucasian
	3 = Latino
	4 = Other

Variable:	Age
Definition:	Predictor: Age
Type:	Continuous

Variable:	Income
Definition:	Predictor: Annual gross income (in dollars)
Type:	Continuous

Variable:	Cigarettes
Definition:	Predictor: Baseline mean number of cigarettes smoked daily
Type:	Continuous

Variable:	Smoking_status
Definition:	Outcome: Smoking status at conclusion of smoking cessation intervention
Type:	Categorical
	0 = Still smoking
	1 = Quit smoking [← *BASIS FOR MODEL*]

This codebook includes six variables: The five predictor variables consist of two categorical variables (*Gender* and *Race*) and three continuous variables (*Age, Income,* and *Cigarettes*). The (one) outcome variable is a dichotomous categorical variable (*Smoking_status*).

For the most part, this codebook resembles the others presented throughout this text; in fact, there are no modifications to the way that the continuous variables (*Age, Income,* and *Cigarettes*) are presented, but in preparation for logistic regression processing, notice that some of the attributes for the categorical variables are different:

- The values for each categorical variable are arranged vertically to facilitate better visual clarity.
- The numbering of the categorical values begins with 0 instead of 1.
- For each of the categorical predictor variables (*Gender* and *Race*), the first category (0) is identified as the *REFERENCE* category; this will be explained in further detail in the Results section.
- For the outcome variable (*Smoking_status*), the last category (1 = *Quit smoking*) is identified as the *BASIS* for this logistic regression model; this will be explained in further detail in the Results section.

☑ Pretest Checklist

Logistic Regression Pretest Checklist

☑ 1. *n* quota*

☑ 2. Normality*

☑ 3. Multicollinearity*

*Run prior to logistic regression test

Three pretest criteria need to be assessed to better ensure the robustness of the findings: **(1) *n* quota, (2) normality,** and **(3) multicollinearity.**

☑ Pretest Checklist Criterion 1—*n* Quota

Considering that the logistic regression statistic is unique in that it accommodates both continuous and categorical predictor variables, there are several steps involved in determining the minimum required sample size.

First, determine the minimum *n:*

1. Count the total number of continuous predictor variables (*Age, Income, Cigarettes*) = **3**.

2. Count the number of categories contained within each categorical variable (*Gender* and *Race*) and subtract 1 from each:

   - *Gender* has 2 categories (*Female*, *Male*): 2 − 1 = **1.**
   - *Race* has 5 categories (*African American, Asian, Caucasian, Latino, Other*): 5 − 1 = **4.**

3. Add the (**bold**) figures together: **3 + 1 + 4 = 8.**

4. Multiply that sum by 10: **8** × 10 = 80. The minimum *n* required to run this logistic regression is 80.

You may find it clearer to organize the variables in a table (Table 13.1).

- For each continuous variable, *n* = 10.
- For each categorical variable, *n* = (number of categories − 1) × 10.

Table 13.1   Assess the Variables to Determine the Minimum *n* Required to Run a Robust Logistic Regression (in This Case, *n* = 80).

Variable	Type	Categorical (Categories − 1) × 10	Continuous 10
Age	Continuous		10
Income	Continuous		10
Cigarettes	Continuous		10
Gender	Categorical	10	
Race	Categorical	40	
Total *n* quota = **80**		**50**	**30**

Proceed by verifying that the data set contains the minimum required *n* (80):

5. On the SPSS main menu, click on *Analyze, Descriptive Statistics, Frequency* (Figure 13.2).
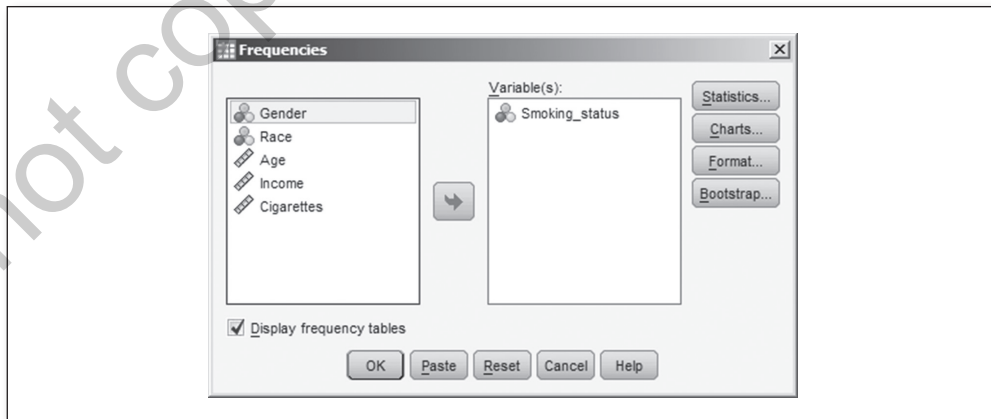
Figure 13.2 To determine the *n* of the data set, click on *Analyze, Descriptive Statistics, Frequencies*.
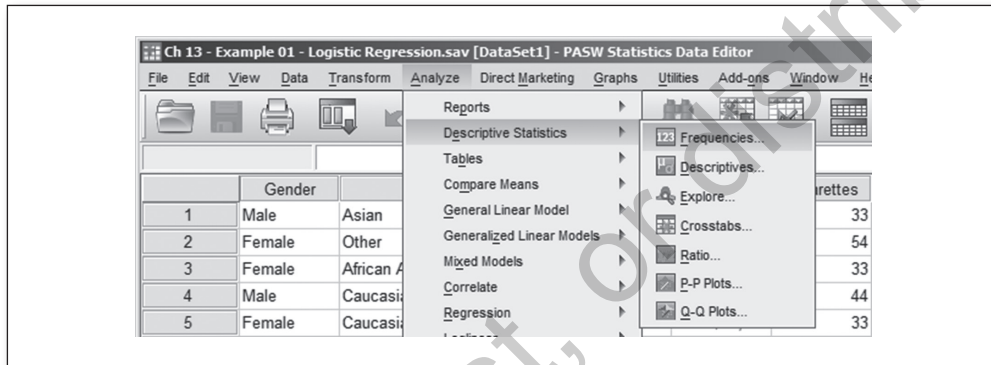
6. Move the outcome variable (*Smoking_Status*) into the *Variable(s)* window (Figure 13.3).

Figure 13.3 On the *Frequencies* menu, move the outcome variable (*Smoking_status*) into the *Variable(s)* window.

7. Click on *OK*.

The *Smoking_status* table shows a *Total Frequency* (*n*) of 218, which is greater than the minimum required (*n* = 80); hence, this pretest criterion is satisfied (Table 13.2).

**Table 13.2**   Descriptive Statistics for Smoking Status: Total (*n*) = 218.

Smoking_status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Still smoking	111	50.9	50.9	50.9
	Quit smoking	107	49.1	49.1	100.0
	Total	218	100.0	100.0	

☑ Pretest Checklist Criterion 2—Normality

Each of the (three) continuous variables should be normally distributed. This will involve ordering histograms with normal curves and inspecting each for normality. The procedure for ordering these charts is detailed on page **58**; at the ★ icon, move *Age, Income, Cigarettes* into the *Variable(s)* window (Figure 13.4).
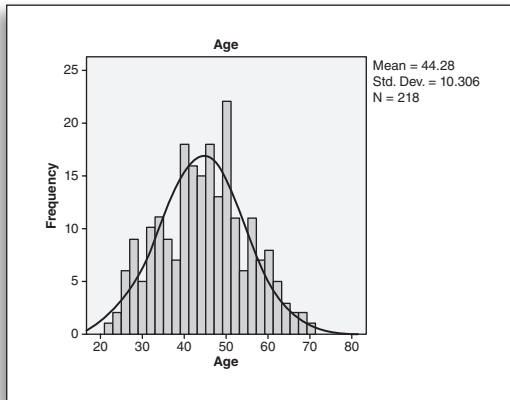
Figure 13.4   To order histograms with normal curves for the continuous variables, click on *Analyze, Descriptive Statistics, Frequencies*.
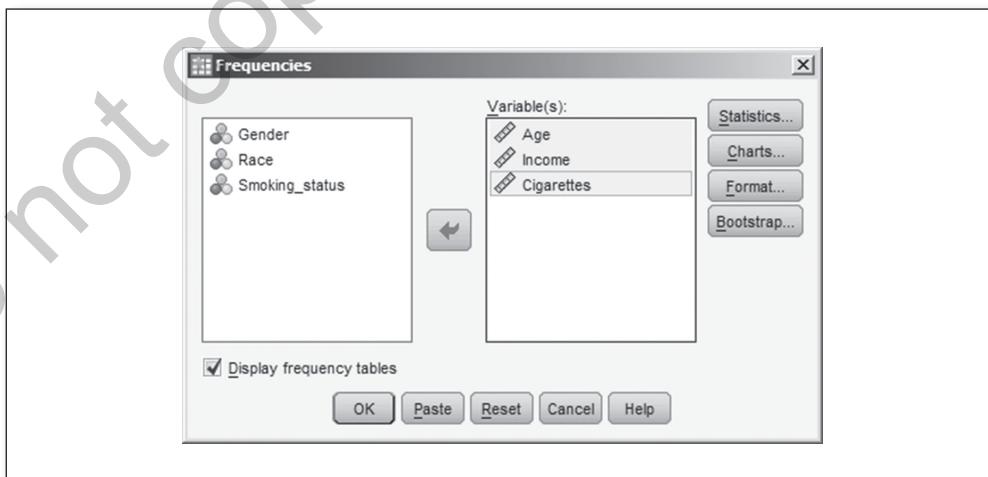
The histograms with normal curves for *Age, Income,* and *Cigarettes* (Figures 13.5, 13.6, and 13.7) are normally distributed; hence, this criterion is satisfied.

Figure 13.5	Histogram for *Age*.

Mean = 44.28
Std. Dev. = 10.306
N = 218

Figure 13.6	Histogram for *Income*.

Mean = $72,899.08
Std. Dev. = $19,271.978
N = 218
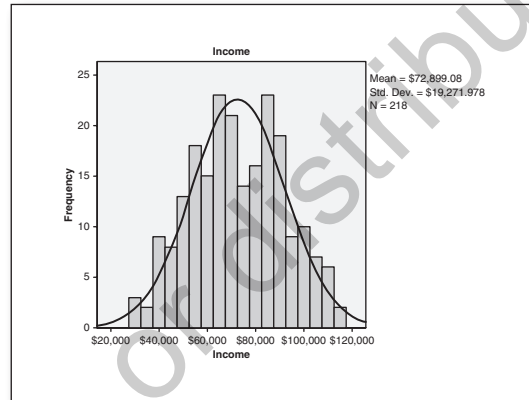
Figure 13.7	Histogram for *Cigarettes*.

Mean = 33.89
Std. Dev. = 9.875
N = 218
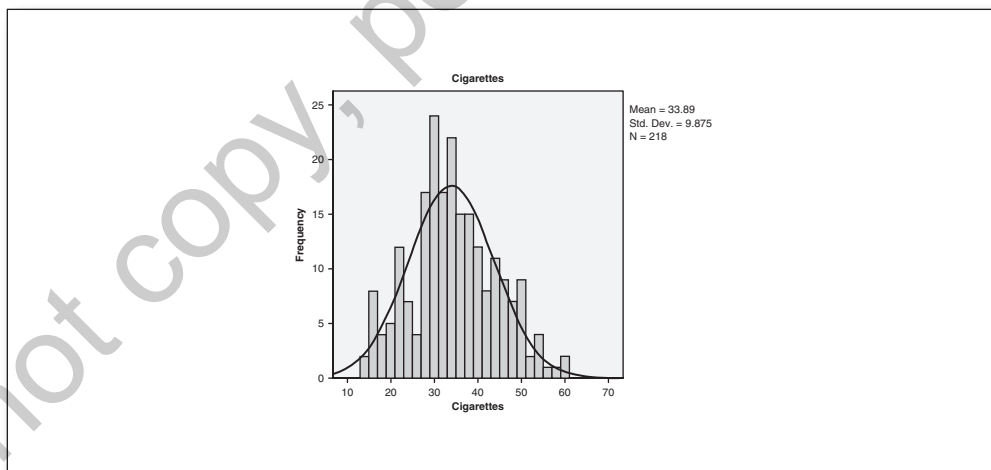
Pretest Checklist Criterion 3—Multicollinearity

The term *multicollinearity* describes two continuous variables that are very highly correlated. Loading two such variables into a logistic regression model

essentially constitutes double-loading the processor; *checking that we do not have multicollinearity* assures us that each continuous variable that we intend to load into the logistic regression model is (statistically) unique. As a rule, variables that have a (Pearson) correlation that is either less than −.9 or greater than +.9 are considered too highly correlated, which would constitute multicollinearity. In such instances, one of the variables should be eliminated from the model—presumably the one that has less utility (e.g., conceptually less critical, more costly/inconvenient to gather). We will use the ±.9 cutoff to assess for multicollinearity, but this threshold is not set in stone; some statisticians set the cutoff at ±.7 or ±.8.

It is possible to construct a perfectly viable logistic regression model primarily consisting of categorical variables. If there are 0 or 1 continuous predictor variables in the logistic regression model, then you do not need to be concerned with multicollinearity—there would be no (other) continuous variable(s) to be too highly correlated with. In such cases, you can simply skip this step.

Considering that there are three continuous predictor variables in this model, we need to check for multicollinearity; we will run a correlational analysis involving all (three) continuous variables:

1. On the main screen, click on *Analyze, Correlate, Bivariate* (Figure 13.8).
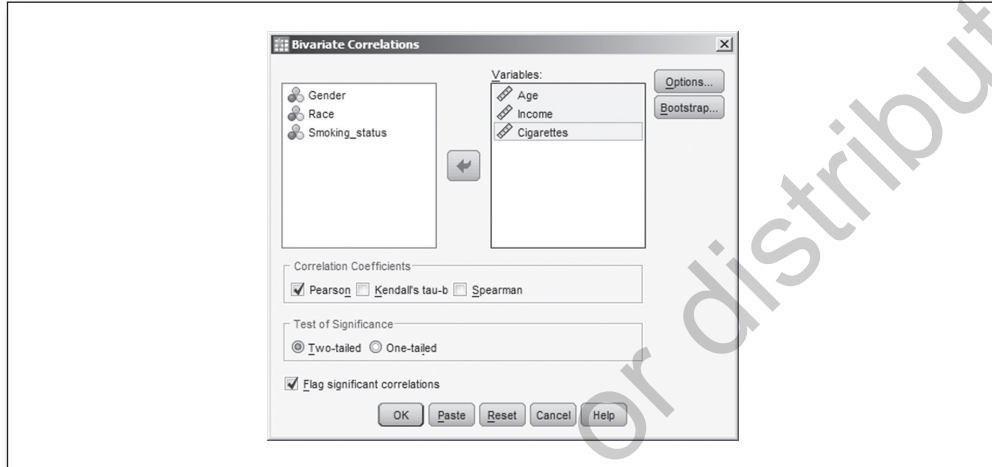
Figure 13.8 Check for multicollinearity; click on *Analyze, Correlate, Bivariate.*
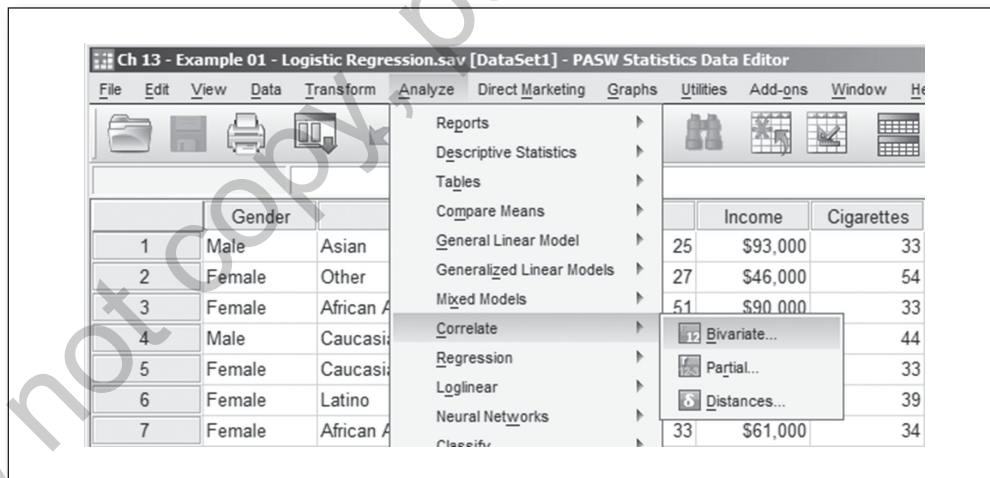
2. On the *Bivariate Correlations* menu (Figure 13.9), move the continuous variables (*Age, Income, Cigarettes*) into the *Variables* window.

3. Click on *OK.*

Bivariate Correlation menu—load *Age, Income,* and *Cigarettes* into *Variables* window.

The *Correlations* table indicates the correlations between each pair of continuous variables (Table 13.3). For further clarity, these correlations are summarized in Table 13.4.

Table 13.3 *Correlations* Table Shows the Pearson Correlations for Each Pair of Continuous Variables (Age : Income, Age : Cigarettes, Income : Cigarettes).

Correlations

		Age	Income	Cigarettes
Age	Pearson Correlation	1	.073	-.250**
	Sig. (2-tailed)		.284	.000
	N	218	218	218
Income	Pearson Correlation	.073	1	-.095
	Sig. (2-tailed)	.284		.161
	N	218	218	218
Cigarettes	Pearson Correlation	-.250**	-.095	1
	Sig. (2-tailed)	.000	.161	
	N	218	218	218

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

Table 13.4 Summary Correlation Table.

Pair	Pearson
Age : Income	.073
Age : Cigarettes	−.250
Income : Cigarettes	−.095

Each of the Pearson correlation scores are between −.9 and +.9; hence, this criterion is satisfied.

Test Run

1. On the main SPSS menu, click on *Analyze, Regression, Binary Logistic* (Figure 13.10).

Figure 13.10 To run a logistic regression, click on *Analyze, Regression, Binary Logistic*.
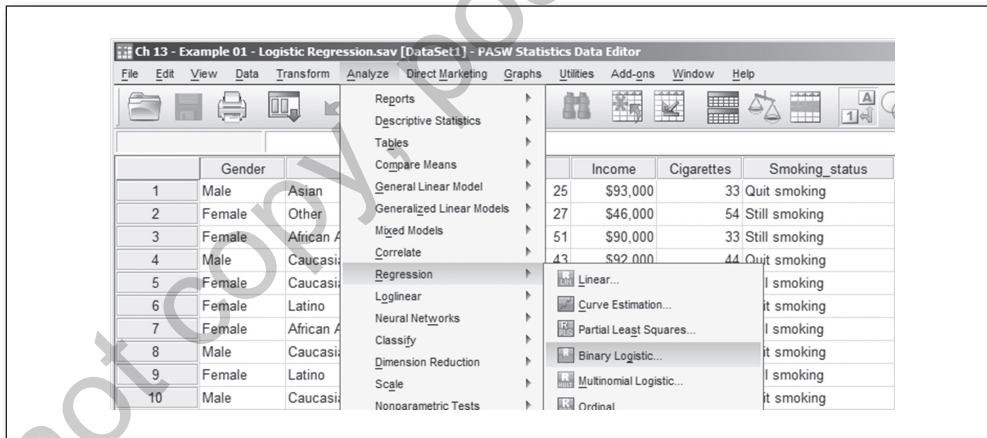
2. On the *Logistic Regression* menu (Figure 13.11), move the outcome variable *(Smoking_status)* into the *Dependent box*.

3. Move the predictor variables (*Gender, Race, Age, Income, Cigarettes*) into the *Covariates* window.

Figure 13.11 *Logistic Regression* menu—move outcome variable into the *Dependent* box and predictor variables into the *Covariates* window.
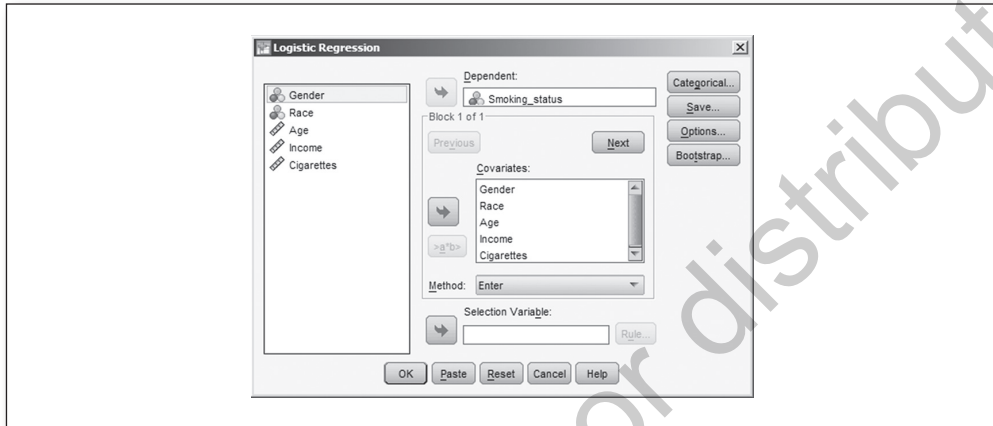
4. Next, identify the categorical variables: Click on *Categorical*.

5. On the *Logistic Regression: Define Categorical Variables* menu, move the two categorical variables (*Gender* and *Race*) into the *Categorical Covariates* window (Figure 13.12).

Figure 13.12 *Logistic Regression: Define Categorical Variables* menu—move categorical predictor variables into the *Categorical Variables* window.
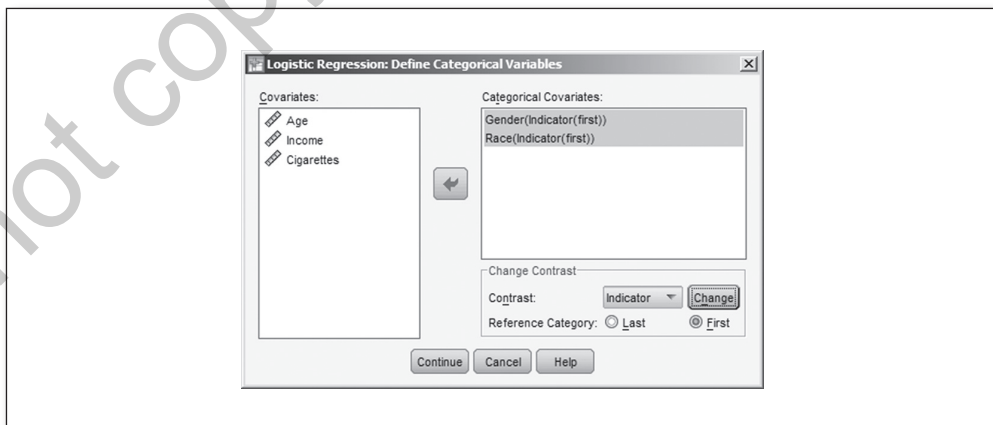
6. In this example, the first category within each categorical variable will be designated as the *Reference Category;* as such, select (highlight) the two variables in the *Categorical Covariates* window.

7. For the *Reference Category,* click on ⊙ *First,* and click on *Change.*

NOTE: The notion of the *Reference Category* will be discussed in the Results section.

8. Click on *Continue*—this will return you to the *Logistic Regression* menu.

9. Click on *Options.*

10. On the *Logistic Regression Options* menu, check ☑ *CI for exp*(*B*). Use the default value of 95% (Figure 13.13).

Figure 13.13 *Logistic Regression Options* menu—select the *CI for exp(B)* (confidence interval), using the default value of 95%.
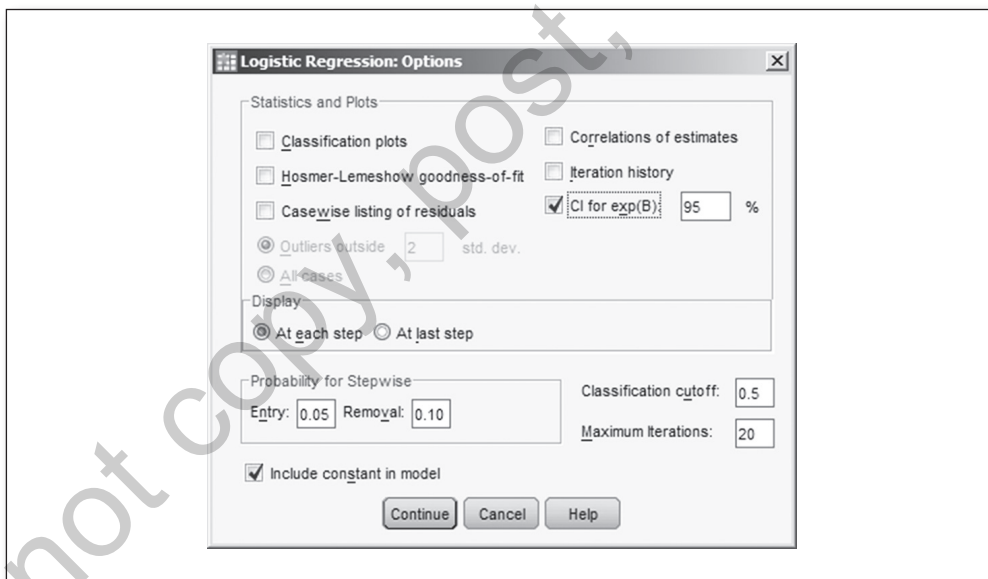
11. Click on *Continue*—this will return you to the *Logistic Regression* menu.

12. Click on *OK.*

▬ Results

Examine the first row of the *Omnibus Tests of Model Coefficients* table (Table 13.5); the Sig. (*p*) is .000, which is less than .05; this indicates that somewhere in the model, at least one of the predictor variables is statistically significant with respect to predicting the outcome variable (did/did not quit smoking). If this Sig. (*p*) is greater than .05, then this would indicate that the overall model is statistically insignificant, meaning that none of the predictors strongly predict the outcome variable. To discover *which* predictor variable(s) statistically significantly predict the outcome variable, we will look to the *Variables in the Equation* table and identify the rows where Sig. (*p*) is less than or equal to .05.

Table 13.5 Omnibus Tests of Model Coefficients Table Shows a Sig. (*p*) < .05—Hence, at Least One Predictor in the Model Is Statistically Significant.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	129.192	8	.000
	Block	129.192	8	.000
	Model	129.192	8	.000

Comprehending R^2

Prior to investigating the *Variables in the Equation* table, we will take a short diversion to discuss the R^2 statistic. The R^2 is regularly used in multiple regression, which is different from logistic regression; while both can process continuous and categorical predictors, the outcome variable in multiple regression is *continuous,* whereas the outcome variable in logistic regression is *categorical*—specifically, dichotomous.

The R^2 is used to express the extent to which the predictors account for the variability observed in the outcome variable. For example, suppose a multiple regression model produces R^2 = .321; we would document it as such: *This model accounts for 32.1% of the variability observed in the outcome variable.* The remaining 67.9% (100% − 32.1% = 67.9%) is (statistically) referred to as "error." In this context, the term *error* does not necessarily imply that someone made a mistake; rather, it simply means that while the predictor variables in the model account for 32.1% of the variability observed in the outcome variable, we do not (yet) know what *other* predictors account for the remaining 67.9% of the variability observed in the outcome variable.

Table 13.6 *Model Summary* Table Shows Nagelkerke $R^2 = .596$.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	172.947[a]	.447	.596

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Currently, there is no perfect R^2 equation for logistic regression; hence, this statistic is commonly referred to as a "pseudo-R^2." In Table 13.6, notice that the Cox & Snell $R^2 = .447$, whereas the Nagelkerke $R^2 = .596$; clearly, these results are quite different. Typically, the Nagelkerke R^2 is considered the better option, but there remains some debate regarding the wisdom of reporting the R^2 for logistic regression. If this statistic were to be included in the documentation, it could be phrased as such: *The Nagelkerke* R^2 *indicates that this model accounts for 59.6% of the variability in smoking cessation.*

The essential findings of the logistic regression are found in the *Variables in the Equation* table (Table 13.7).

Table 13.7 Unedited *Variables in the Equation* Table.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Gender(1)	3.101	.483	41.269	1	.000	22.223	8.628	57.241
	Race			9.873	4	.043			
	Race(1)	.913	.917	.990	1	.320	2.492	.413	15.043
	Race(2)	.218	.615	.125	1	.723	1.243	.372	4.150
	Race(3)	.504	.601	.704	1	.402	1.656	.510	5.376
	Race(4)	2.082	.721	8.335	1	.004	8.022	1.951	32.973
	Age	.101	.023	19.505	1	.000	1.107	1.058	1.158
	Income	.000	.000	1.138	1	.286	1.000	1.000	1.000
	Cigarettes	-.056	.023	5.904	1	.015	.946	.904	.989
	Constant	-3.720	1.528	5.926	1	.015	.024		

a. Variable(s) entered on step 1: Gender, Race, Age, Income, Cigarettes.

Notice that the numeric values are presented for the categorical variables but not the assigned text labels. In preparation for the documentation process, it is recommended that you manually include the text value labels for each categorical variable:

1. Copy the *Variables in the Equation* table from SPSS into the word processor.

2. Refer to the codebook (see the ★ icon on page **294**), and manually type in the text value labels that correspond to each categorical variable (you will need to adjust the column sizes of the table).

3. If a separate codebook document is not provided, these categorical labels can be derived from viewing the *Values* assigned to each categorical variable on the *Variable View* screen.

4. The **[BRACKETED BOLD]** text in Table 13.8 was typed in manually.

Table 13.8 Edited *Variables in the Equation* Table.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Gender(1) **[0 = Female, 1 = Male]**	3.101	.483	41.269	1	.000	22.223	8.628	57.241
	Race **[0 = African American]**			9.873	4	.043			
	Race(1) **[1 = Asian]**	.913	.917	.990	1	.320	2.492	.413	15.043
	Race(2) **[2 = Caucasian]**	.218	.615	.125	1	.723	1.243	.372	4.150
	Race(3) **[3 = Latino]**	.504	.601	.704	1	.402	1.656	.510	5.376
	Race(4) **[4 = Other]**	2.082	.721	8.335	1	.004	8.022	1.951	32.973
	Age	.101	.023	19.505	1	.000	1.107	1.058	1.158
	Income	.000	.000	1.138	1	.286	1.000	1.000	1.000
	Cigarettes	-.056	.023	5.904	1	.015	.946	.904	.989
	Constant	-3.720	1.528	5.926	1	.015	.024		

a. Variable(s) entered on step 1: Gender, Race, Age, Income, Cigarettes.

NOTE: **[BRACKETED BOLD]** text manually typed in to clearly label categorical variables.

Notice that the confidence interval [*95% C.I. for EXP(B)*] is included in Table 13.8. The first row (*Gender*) indicates a lower CI of 8.628 and an upper CI of 57.241, pertaining to the Exp(B) of 22.233. This is saying that for the odds ratio pertaining to *Gender,* 95% of the values are expected to be between 8.628 and 57.241. Confidence intervals are traditionally included in logistic regression documentation for statistically significant predictors.

H_0 HYPOTHESIS RESOLUTION

The *Omnibus Tests of Model Coefficients* table indicates a Sig. (*p*) value of .000; since this is less than the .05 α level, this indicates that at least one of the (predictor) variables is statistically significant; hence, we reject H_0 and accept H_1.

REJECT: H_0: Age, income, baseline smoking, gender, and race do not influence one's success in a smoking cessation intervention.

ACCEPT: H_1: Age, income, baseline smoking, gender, and race influence one's success in a smoking cessation intervention.

The next step is to identify and document the specific predictor variable(s) that produced statistically significant results.

DOCUMENTING RESULTS

Documentation Overview

Considering that the logistic regression statistic accommodates an assortment of variables—(1) dichotomous outcome variable, (2) categorical predictor variables, and (3) continuous predictor variables—the documentation procedure will be presented in three parts:

- Part 1: Comprehending the outcome variable
- Part 2: Documenting categorical predictors
- Part 3: Documenting continuous predictors

Additionally, the logistic regression model is versatile in terms of its capacity to produce a variety of results. As such, this documentation section consists of three models:

- Model 1: Initial results
- Model 2: Selective results
- Model 3: Redefining a reference category

Model 1: Initial Results

Documenting Results Part 1: Outcome Variable

Consider this excerpt from the codebook detailing the dichotomous outcome variable:

Outcome variable: Smoking_status

0 = Still smoking (FAILED)

1 = Quit smoking (SUCCEEDED) [← *BASIS FOR MODEL*]

Although it may sound a bit redundant, since the intended goal of this intervention was to have patients successfully *Quit smoking,* we will want to discuss the results in terms of the characteristics of those who successfully *Quit smoking,* as opposed to those who are *Still smoking;* hence, the label *Quit smoking* is assigned a value of 1 in the outcome variable *Smoking_status*. This will serve as the (semantic) basis for this model. As you will see in Parts 2 and 3, the results in the *Variables in the Equation* table pertain to those who *Quit smoking*.

Table 13.9 Labeled *Variables in the Equation* Table, Focusing on Categorical Variables.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Gender(1) **[0 = Female, 1 = Male]**	3.101	.483	41.269	1	.000	22.223	8.628	57.241
	Race **[0 = African American]**			9.873	4	.043			
	Race(1) **[1 = Asian]**	.913	.917	.990	1	.320	2.492	.413	15.043
	Race(2) **[2 = Caucasian]**	.218	.615	.125	1	.723	1.243	.372	4.150
	Race(3) **[3 = Latino]**	.504	.601	.704	1	.402	1.656	.510	5.376
	Race(4) **[4 = Other]**	2.082	.721	8.335	1	.004	8.022	1.951	32.973
	Age	.101	.023	19.505	1	.000	1.107	1.058	1.158
	Income	.000	.000	1.138	1	.286	1.000	1.000	1.000
	Cigarettes	-.056	.023	5.904	1	.015	.946	.904	.989
	Constant	-3.720	1.528	5.926	1	.015	.024		

a. Variable(s) entered on step 1: Gender, Race, Age, Income, Cigarettes.

Documenting Results Part 2: Categorical Predictors

We will begin by documenting the results from all of the values in each categorical variable regardless of the Sig. (*p*) value to thoroughly demonstrate how to translate the data on this table into appropriately written results. After that process, as expected, we will narrow our discussion to only those variables that are statistically significant (where $p \leq .05$).

This model contains two categorical predictor variables: *Gender* and *Race*. We will begin by interpreting and documenting the *Gender* variable.

For *Gender, Female* is coded as 0, establishing it as the *reference category* for *Gender*. As such, all of the results for *Gender* will be expressed as comparisons to *Females*. Referring to the data in the first column (which contains the variable names) and the figures in the *Exp(B)* column, we document the results as such:

- Males have 22.223 times the odds of quitting smoking compared to females (95% CI 8.63, 57.24) (meaning that the men in this study succeeded in quitting smoking significantly more frequently than women).

Alternatively, this could be rephrased as follows:

- The odds of quitting smoking are 22.223 times higher for males compared to females (95% CI 8.63, 57.24).

For *Race,* the categories are arranged alphabetically; hence, *African American* is coded as 0, establishing it as the reference category for *Race*. As such, all of the results for *Race* will be expressed as comparisons to *African Americans*.

- Asians have 2.492 times the odds of quitting smoking compared to African Americans (95% CI.41, 15.04).
- Caucasians have 1.243 times the odds of quitting smoking compared to African Americans (95% CI.37, 4.15).
- Latinos have 1.656 times the odds of quitting smoking compared to African Americans (95% CI.51, 5.38).
- Others have 8.022 times the odds of quitting smoking compared to African Americans (95% CI 1.95, 32.97).

You may include the corresponding p (Sig.) values, flagging those where $p \leq .05$. Alternatively, you may wish to provide detailed discussion of only those categories wherein $p \leq .05$ (*Other*) and briefly mention the others as statistically insignificant (*Asians, Caucasians, Latinos*). These findings will be carried forward when we draft the abstract.

Categorical Documentation Option: Alternate Write-Up if Exp(B) Is Less Than 1

In the above table, *Gender* produced *Exp(B) = 22.223,* which is greater than 1. Suppose instead of *22.223,* it was .456. When *Exp(B)* is less than 1 for a categorical predictor, the semantics of the write-up may seem a bit awkward (see ORIGINAL documentation phrasing below). One option that may help to clarify the documentation is to "flip" the sentence. This involves calculating the reciprocal of *Exp(B),* which is simply **1 ÷ Exp(B)**; this would be **1 ÷ .456 = 2.193,** and swapping the variable labels in the sentence.

- ORIGINAL: *Males have .456 times the odds of quitting smoking compared to females.*
- FLIPPED: *Females have 2.193 times the odds of quitting smoking compared to males.*

Documenting Results Part 3: Continuous Predictors

Next, we will document the results produced by the three continuous predictors (*Age, Income, Cigarettes*) in Table 13.10.

Table 13.10 Labeled *Variables in the Equation* Table, Focusing on Continuous Variables.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Gender(1) **[0 = Female, 1 = Male]**	3.101	.483	41.269	1	.000	22.223	8.628	57.241
	Race **[0 = African American]**			9.873	4	.043			
	Race(1) **[1 = Asian]**	.913	.917	.990	1	.320	2.492	.413	15.043
	Race(2) **[2 = Caucasian]**	.218	.615	.125	1	.723	1.243	.372	4.150
	Race(3) **[3 = Latino]**	.504	.601	.704	1	.402	1.656	.510	5.376
	Race(4) **[4 = Other]**	2.082	.721	8.335	1	.004	8.022	1.951	32.973
	Age	.101	.023	19.505	1	.000	1.107	1.058	1.158
	Income	.000	.000	1.138	1	.286	1.000	1.000	1.000
	Cigarettes	-.056	.023	5.904	1	.015	.946	.904	.989
	Constant	-3.720	1.528	5.926	1	.015	.024		

a. Variable(s) entered on step 1: Gender, Race, Age, Income, Cigarettes.

Continuous variables are best expressed in terms of odds percentages. There are two possible outcomes and documentation procedures for continuous predictors: Either *Exp(B)* is less than 1 or *Exp(B)* is greater than 1:

Documenting Continuous Variable Results if Exp(B) Is Less Than 1

If Exp(B) < 1, then the percentage = (1 − Exp(B)) × 100.

For *Cigarettes,* Exp(B) = .946; since this is less than 1, this indicates a *decrease*. We compute (1 − .946) × 100 = 5.4; hence, the write-up would be as follows:

- *For every additional cigarette smoked per day, the odds of quitting smoking decreases by 5.4%* (95% CI .90, .99).

Documenting Continuous Variable Results if Exp(B) Is Greater Than 1

If Exp(B) > 1, then the percentage = (Exp(B) − 1) × 100.

For *Age, Exp(B)* = 1.107; since this is greater than 1, this indicates an *increase*. We compute (1.107 − 1) × 100 = 10.7; hence, the write-up would be as follows:

- *For every additional year of age, the odds of quitting smoking increases by 10.7% (95% CI 1.06, 1.16).*

Documenting Continuous Variable Results if Exp(B) Equals 1

For *Income, Exp(B)* = 1.000, indicating that *Income* produced 1:1 odds in terms of Income predicting the likelihood that a participant will quit smoking. In other words, about the same number of people with high income as low income quit smoking—the odds of quitting smoking are the same regardless of high/low income. As expected, the Sig. (*p*) value for *Income* is .286; since this is greater than the .05 α level, *Income* is considered statistically insignificant when it comes to predicting the likelihood that a participant will successfully quit smoking.

Logistic Regression Documentation Summary

Categorical Predictors

If Exp(B) > 1, then odds ratio = Exp(B).
The comparison category is **more likely** than the reference category to predict the outcome variable (basis for model).

If Exp(B) < 1, then odds ratio can be expressed as 1 ÷ Exp(B) and flip the variables.
The comparison category is **less likely** than the reference category to predict the outcome variable (basis for model).

Continuous Predictors

If Exp(B) > 1, then percentage = (Exp(B) − 1) × 100.
The comparison category **increases** in relation to the outcome variable (basis for model).

If Exp(B) < 1, then percentage = (1 − Exp(B)) × 100.
The comparison category **decreases** in relation to the outcome variable (basis for model).

Abstract for Model 1: Initial Results

You may include the corresponding *p* (Sig.) values, flagging those where $p \leq .05$. Alternatively, you may wish to provide detailed discussion of only those categories wherein $p \leq .05$ and briefly mention the others as statistically insignificant:

As a public service, the Acme Health Center advertises and offers a free 90-day smoking cessation program, consisting of nurse-facilitated psychoeducational meetings, peer support from those who have been smoke free for more than 1 year, and multimedia resources designed to promote smoking cessation.

*At the conclusion of the intervention, each participant (*n *= 218) responded to a self-administered anonymous Smoking Cessation Survey card, which gathered*

data on gender, race, age, income, baseline mean number of cigarettes smoked per day, and current smoking status (still smoking/quit smoking).

To better comprehend the factors associated with successfully quitting smoking, we conducted a logistic regression analysis. We discovered that males had 22.223 times the odds of quitting smoking compared to females (p < .001) (95% CI 8.63, 57.24). Those who indicated that their race designation was "Other" had 8.022 times the odds of quitting smoking compared to African Americans (p = .004) (95% CI 1.95, 32.97). Older participants were more likely to quit than those who were younger; for every additional year of age, the odds of quitting smoking increased by 10.7% (p < .001) (95% CI 1.06, 1.16). We also discovered that baseline smoking was an influential factor; for every additional cigarette smoked per day, the odds of quitting smoking decreased by 5.4% (p = .015) (95% CI .90, .99). Income was not found to be a viable predictor when it comes to predicting who successfully quit smoking (p = .286) (95% CI 1.00, 1.00).

Considering the volume of results produced by a logistic regression analysis in light of the relative brevity of an abstract (usually about 200 words), not every possible statistic was discussed. In a more comprehensive Results section, other statistical findings could be included (e.g., overall percentage of those who quit smoking, descriptive statistics for each categorical and continuous variable, discussion of statistically insignificant predictors).

Model 2: Selective Results

In the initial model, the results reflected the overall findings from both *Genders* (*Female* and *Male*); it was found that the men in this group were substantially more successful in quitting smoking than women. Such an observation may lead one to ponder: *Among the males (only), what were the significant predictors when it comes to successfully quitting smoking?* Statistically, this question is asking, *What would the results of this logistic regression look like if the females were removed from the picture?* This would be akin to recruiting only males to partake in this intervention.

Fortunately, we do not need to repeat this study as a *men-only* intervention to address this question; instead, we can access the existing database, using the *Select Cases* function to process only those records (rows of data) pertaining to males (*Gender* = 1).

1. On the main screen, click on the *Select Cases* icon.

2. On the *Select Cases* menu, click on ⊙ *If condition is satisfied*.

3. Click on *If.*

4. On the *Select Cases: If* menu, enter *Gender = 1* (Figure 13.14).

Figure 13.14   Selecting cases where *Gender* = 1 (Male).
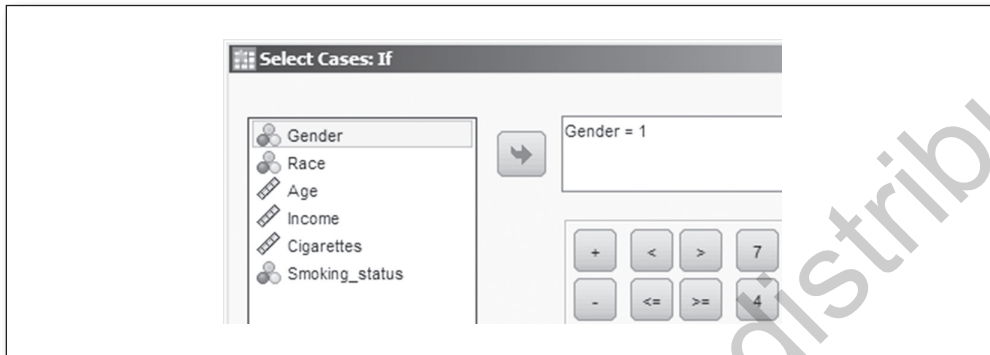
5. Click on *Continue* (this will return you to the *Select Cases* menu).

6. On the *Select Cases* menu, click on *OK*.

7. Now that only *Males* are selected, proceed to rerun the logistic regression analysis using each of the steps detailed in this chapter, as if all of the records were in play.

This analysis produces a table (Table 13.11) pertaining to the data gathered from the *Males* only. As expected, these findings are quite different compared to the initial run, which involved both genders.

Table 13.11   Labeled *Variables in the Equation* Table, for *Males*.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1ᵃ	Race [0 = African American]			5.976	3	.113			
	Race(1) [1 = Asian]	2.395	13243.109	.000	1	1.000	10.967	.000	.
	Race(2) [2 = Caucasian]	-17.739	8095.370	.000	1	.998	.000	.000	.
	Race(3) [3 = Latino]	-20.050	8095.370	.000	1	.998	.000	.000	.
	Age	.163	.057	8.053	1	.005	1.176	1.052	1.316
	Income	.000	.000	.330	1	.566	1.000	1.000	1.000
	Cigarettes	-.151	.053	8.147	1	.004	.860	.775	.954
	Constant	18.863	8095.371	.000	1	.998	1.557E8		

a. Variable(s) entered on step 1: Race, Age, Income, Cigarettes.

First, notice that even if you attempted to load *Gender* as a variable, it was eliminated from the process since (now) *Gender* contains only one (selected) value: 1, signifying *Males* (only). Since all of the values for *Gender* = 1, technically, *Gender* ceases to be a *variable* because it does not vary; it is constantly 1, and hence, it is considered a constant. In this process, a constant has no predictive capacity since it is constantly 1 no matter what is happening among the other variables. As such, *Gender* is appropriately eliminated from the table.

Next, notice that *Race(4)* is missing. This is because there are no *Males* who specified their *Race* as category 4 (*Other*). For proof, run descriptive statistics (with a bar chart) for the *Race* variable, and notice that *Other* is absent.

As for documenting the results, although we may notice that *Asians* have 10.967 times the odds of quitting smoking compared to *African Americans,* we see that this finding is considered statistically insignificant (*p* = 1.000).

Abstract for Model 2: Selective Results

*Assessing only the male participants, we discovered that those who were older were more likely to quit than those who were younger; for every additional year of age, the odds of quitting smoking increased by 17.6% (*p* = .005) (95% CI 1.05, 1.32). We also discovered that baseline smoking was an influential factor; for every additional cigarette smoked per day, the odds of quitting smoking decreased by 14% (*p* = .004) (95% CI .77, .95).*

Model 3: Redefining a Reference Category

Data Set

Use the following data set: **Ch 13 – Example 02 – Logistic Regression.sav.**

In the prior two models, the reference category for *Race* has been *African American,* which produced statistics that compared all of the other racial categories to *African Americans* in terms of quitting smoking. This designation was merely due to the alphabetical arrangement of the categories—*African Americans* just happen to occupy the

Table 13.12	To Set *Other* as the New Reference Category, Recode *Other* From 4 to 0 and Recode *African American* From 0 to 4.

Recoding Race Reference (Swap **African American** With **Other**)	
Race	**Race_0**
0 = **African American [← *REFERENCE*]** 1 = **Asian** 2 = Caucasian 3 = Latino **4** = **Other**	**0** = **Other [← *REFERENCE*]** 1 = Asian 2 = Caucasian 3 = Latino **4** = **African American**

first position. There may be instances where you want to designate a different category as the reference category for a variable. For this example, we will change the original coding so that *Other* becomes the new reference category for *Race*. One way to do this involves changing the categorical coding of *Other* from 4 to 0 and recode *African American* from 0 to 4.

The current database (**Ch 13 – Example 02 – Logistic Regression.sav**) is the same as the original, except the variable *Race* (which was initially coded as 0 = *African American* and 4 = *Other*) has been recoded to create *Race_O,* wherein 0 = *Other* and 4 = *African American* (Table 3.12). With *Other* now serving as the reference category for *Race_O,* all racial results will be presented as comparisons to *Other*.

As a side note, the Recode process was used to create *Race_O,* wherein all of the 0s were replaced with 4s, all of the 4s were replaced with 0s, and all of the other numbers (1, 2, 3) were kept as is. Finally, the value labels for *Race_O* were (manually) edited accordingly: 0 was changed to *Other,* and 4 was changed to *African American*.

The step-by-step instructions for this recoding procedure are detailed in Chapter 14, on page **343** at the ★ icon.

Table 13.13   Labeled *Variables in the Equation* Table, With *Other* Set to Reference Category in *Race_O.*

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Gender(1) **[0 = Female, 1 = Male]**	3.101	.483	41.269	1	.000	22.223	8.628	57.241
	Race_O **[0 = Other]**			9.873	4	.043			
	Race_O(1) **[1 = Asian]**	-1.169	.956	1.495	1	.221	.311	.048	2.024
	Race_O(2) **[2 = Caucasian]**	-1.864	.716	6.779	1	.009	.155	.038	.631
	Race_O(3) **[3 = Latino]**	-1.578	.624	6.399	1	.011	.206	.061	.701
	Race_O(4) **[4 = African American]**	-2.082	.721	8.335	1	.004	.125	.030	.512
	Age	.101	.023	19.505	1	.000	1.107	1.058	1.158
	Income	.000	.000	1.138	1	.286	1.000	1.000	1.000
	Cigarettes	-.056	.023	5.904	1	.015	.946	.904	.989
	Constant	-1.637	1.446	1.283	1	.257	.194		

a. Variable(s) entered on step 1: Gender, Race_O, Age, Income, Cigarettes.

Notice that the figures for *Race* are different from the initial results, since the reference category is now *Other* instead of *African American*. Additionally, notice that all of the other figures match the results in the initial run. The point is, recoding a variable only changes the presentation of the data contained within that variable.

As expected, a new write-up for *Race_O* is warranted:

Abstract for Model 3: Redefining a Reference Category

In terms of race, Caucasians have .155 times the odds of quitting smoking compared to those who identify their race as Other (p = .009) (95% CI.04, .63). Latinos were found to have .206 times the odds of quitting smoking compared to those who identify their race as Other (p = .011) (95% CI .06, .70), and African Americans have .125 times the odds of quitting compared to Other (p = .015) (95% CI .03, .51).

Alternatively, since these odds ratios are less than 1, the reading may be clearer if the semantics were flipped and the reciprocals [1 ÷ Exp(B)] were presented:

Those who identified their race as Other had 6.45 times the odds of quitting smoking compared to Caucasians (p = .009) (95% CI .04, .63), and Other (race category) had 4.85 times the odds of quitting smoking compared to Latinos (p = .011) (95% CI .06, .70). Additionally, Other had 8 times the odds of quitting smoking compared to African Americans (p = .015) (95% CI .03, .51).

Multiple Regression (R^2)—An Overview

Whereas logistic regression provides odds ratios to determine the influences that continuous and categorical predictors have on a *dichotomous* outcome variable, multiple regression (R^2) can be used to comprehend the influences of predictor variables when the outcome variable is *continuous.*

In the example processed in this chapter, the outcome variable, *Smoking_status,* is a dichotomous variable; the two categories are *Still smoking* and *Quit smoking,* but suppose instead of this dichotomous coding, *Smoking_status* had been coded as a continuous variable, wherein *Smoking_status* would now contain the average number of cigarettes that each participant smokes per day at the conclusion of the intervention. In this case, 0 would indicate that the participant has successfully quit smoking. Whereas logistic regression produces results in the form of odds ratios, the multiple regression identifies the predictors that have a statistically significant correlation to the outcome variable and expresses the results in terms of percentages. The following is a sample of how such results would be documented:

We recruited 218 smokers to participate in a smoking cessation intervention. Initially, we gathered data detailing each participant's gender, race, age, income, and average number of cigarettes smoked per day. At the conclusion of the treatment, we measured the (new) average number of cigarettes smoked per day (0 = quit smoking).

Multiple regression analysis revealed an overall R^2 of .40 (α = .05), wherein gender accounts for 30% of the variability observed in the outcome variable (posttreatment smoking), race accounts for an additional 7%, and baseline smoking rate accounts for 3%.

In this (hypothetical) example, notice that *age* and *income* are not mentioned; this is because the regression processor determined that they are not statistically significantly correlated to the outcome variable (posttreatment smoking). Also notice that the overall R^2 only accounts for 40% of the variability observed in the outcome variable, so the

question stands: *What about the other 60% (100 − 40 = 60)?* The answer to that is *error*. In this context, "error" does not imply that somebody made a mistake; rather, this model is saying that three of the predictor variables (gender, race, and baseline smoking rate) account for 40% of the variability observed in the outcome variable, leaving 60% unaccounted for. Basically, this is saying that other predictors pertain to the outcome variable, which are not included in this model. If this study were to be repeated, we might consider retaining the three statistically significant predictors (gender, race, and baseline smoking rate), dropping the statistically insignificant predictors (age and income), and include some other, hopefully more relevant, predictor variables to increase the overall R^2.

GOOD COMMON SENSE

Logistic regression is a sophisticated type of analytic procedure that enables one to gain a deeper understanding of the relationships among the variables in terms of predicting a dichotomous outcome. In some cases, the findings from a logistic regression model can be used to predict/anticipate the likelihood of an outcome.

Despite the detailed findings produced by logistic regression, keep in mind that the model pertains to *a group of people*—it does not describe or predict the outcome of any one individual. In the same way that descriptive statistics can be used to compute the mean age of people in a sample, knowing that mean age (e.g., 25) does not empower you to point to any one person in the sample (or population) and confidently proclaim, "You are 25 years old." Keep in mind that this same principle also applies to more advanced processes, such as logistic regression.

Key Concepts

- Logistic regression
- Pretest checklist:
  - Sample size
  - Normality
  - Multicollinearity
- R^2 statistic

- Categorical variable labeling
- Selectively processing
- Categorical recoding principles
- Hypothesis resolution
- Documenting results
- Multiple regression overview
- Good common sense

Practice Exercises

Exercise 13.1

A public health nurse has conducted a survey of people in the community to better comprehend the effectiveness of the flu shot this season using the following survey instrument:

Flu Survey

1. Gender: ☐ Female ☐ Male

2. How old are you? _____

3. Did you have a flu shot this season? ☐ No ☐ Yes

4. Do you have any chronic disease(s)? ☐ No ☐ Yes

5. Have you been sick with the flu this season? ☐ No ☐ Yes

Data set: **Ch 13 – Exercise 01A.sav**

Codebook

Variable:	Flu_sick
Definition:	Outcome variable: Did this person get sick with the flu this season?
Type:	Categorical (0 = Got the flu, 1 = No flu)

Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)

Variable:	Flu_shot
Definition:	Predictor variable: Did person have a flu shot this season?
Type:	Categorical (0 = Got a flu shot, 1 = Did not get a flu shot)

Variable:	Chronic_disease
Definition:	Predictor variable: Does the person have chronic disease(s)?
Type:	Categorical (0 = Has chronic disease(s), 1 = No chronic disease(s))

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 01B.sav.**

NOTE: This data set (**Ch 13 – Exercise 01B.sav**) is the same as the first data set except the *Age* variable has been recoded from a continuous variable that contained the actual ages to a categorical variable, now coded as Pediatric/Adult, using the following recoding criteria:

- If Age < 18, then recode as 0 = Pediatric
- If Age ≥ 18, then recode as 1 = Adult

The corresponding modification has been made to the codebook:

Variable: Age
Definition: Predictor variable: Age
Type: Categorical (0 = Pediatric, 1 = Adult)

Exercise 13.2

To better comprehend the characteristics of patients who may be susceptible to developing a rash from Drug A, a nurse gathers data on patients who have been prescribed this medication. Since it is known that taking Drug B, which is commonly prescribed with Drug A, may cause complications, the study enquires about Drug B usage. Additional variables of interest include the duration of the dosage and the patient's age.

Data set: **Ch 13 – Exercise 02A.sav**

Codebook

Variable: Rash
Definition: Outcome variable: Did the patient develop a rash?
Type: Categorical (0 = No rash, 1 = Rash)

Variable: Drug_B
Definition: Predictor variable: Is the patient also taking Drug_B?
Type: Categorical (0 = Taking Drug B, 1 = Not taking Drug B)

Variable: Duration
Definition: Predictor variable: Number of days the patient has been taking Drug A
Type: Continuous

Variable: Age
Definition: Predictor variable: Age
Type: Continuous

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

    c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

    d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 02B.sav.**

Exercise 13.3

A nurse on the quality improvement team has gathered data from three surgical facilities—Northview Surgical Center, South Hills Hospital, and Central Health Clinic—to determine the factors associated with surgical complications.

Data set: **Ch 13 – Exercise 03A.sav**

Codebook

Variable:	Complications
Definition:	Outcome variable: Did the patient experience surgical complication(s)?
Type:	Categorical (0 = No surgical complications, 1 = Surgical complications)

Variable:	Inpatient
Definition:	Predictor variable: Was the surgery inpatient?
Type:	Categorical (0 = Inpatient, 1 = Outpatient)

Variable:	Facility
Definition:	Predictor variable: Surgical facility
Type:	Categorical (0 = Northview Surgical Center, 1 = South Hills Hospital, 2 = Central Health Clinic)

Variable:	Laparoscopic
Definition:	Predictor variable: Laparoscopic surgery
Type:	Categorical (0 = Laparoscopic, 1 = Not laparoscopic)

    a. Write the hypotheses.

    b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

    c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

    d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 03B.sav.**

Exercise 13.4

A nurse who oversees emergency training is investigating the factors pertaining to resuscitation survival at Acme Hospital.

Data set: **Ch 13 – Exercise 04A.sav**

Codebook

Variable:	Survival
Definition:	Outcome variable: Did the patient survive resuscitation?
Type:	Categorical (0 = Did not survive, 1 = Survived)
Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)
Variable:	Race
Definition:	Predictor variable: Race
Type:	Categorical (0 = African American, 1 = Asian, 2 = Caucasian, 3 = Latino, 4 = Other)
Variable:	Prior_ resuscitation
Definition:	Predictor variable: Did the patient require prior resuscitation during this hospitalization?
Type:	Categorical (0 = No, 1 = Yes)
Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous
Variable:	LOS
Definition:	Predictor variable: Length of stay
Type:	Continuous

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 04B.sav.**

Exercise 13.5

The charge nurse at a dialysis center wants to investigate the characteristics of patients who missed an appointment last month to help focus treatment adherence strategies. In addition to data from patient's records, the nurse administers the Acme Depression Scale to each patient.

Data set: **Ch 13 – Exercise 05A.sav**

Codebook

Variable:	Appointments
Definition:	Outcome variable: Did the patient keep all dialysis appointments last month?
Type:	Categorical (0 = Kept all appointments, 1 = Did not keep all appointments)
Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)
Variable:	Race
Definition:	Predictor variable: Race
Type:	Categorical (0 = African American, 1 = Asian, 2 = Caucasian, 3 = Latino, 4 = Other)
Variable:	SES
Definition:	Predictor variable: Socioeconomic status
Type:	Categorical (0 = Lower class, 1 = Middle class, 2 = Upper class)
Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous
Variable:	Dialysis_time
Definition:	Predictor variable: Total length of time on dialysis (in months)
Type:	Continuous
Variable:	Depression
Definition:	Predictor variable: Score on Acme Depression Scale
Type:	Continuous (0 = Low depression . . . 30 = High depression)

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 05B.sav.**

Exercise 13.6

A nurse on the palliative care team wants to better comprehend the attributes of those who opt for hospice care when it is clinically offered and those who do not.

Data set: **Ch 13 – Exercise 06A.sav**

Codebook

Variable:	Palliative_care
Definition:	Outcome variable: Did the patient opt for hospice care?
Type:	Categorical (0 = Refused hospice, 1 = Accepted hospice)

Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)

Variable:	Race
Definition:	Predictor variable: Race
Type:	Categorical (0 = African American, 1 = Asian, 2 = Caucasian, 3 = Latino, 4 = Other)

Variable:	Disease
Definition:	Predictor variable: Primary diagnosis
Type:	Categorical (0 = AIDS, 1 = Cancer, 2 = Cardiac, 3 = Dementia, 4 = Pulmonary, 5 = Stroke, 6 = Other)

Variable:	Religion
Definition:	Predictor variable: Religion
Type:	Categorical (0 = Atheist, 1 = Buddhist, 2 = Catholic, 3 = Hindu, 4 = Jewish, 5 = Other)

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 06B.sav.**

NOTE: This data set is the same as the first one, but the following variables have been recoded as such (*Cancer* is now the reference category for the *Disease* variable):

Variable:	Disease
Definition:	Predictor variable: Primary diagnosis
Type:	Categorical (0 = Cancer, 1 = AIDS, 2 = Cardiac, 3 = Dementia, 4 = Pulmonary, 5 = Stroke, 6 = Other)

Exercise 13.7

A nurse on the Patient Care Committee wants to better comprehend the attributes of those who opt for DNR (do not resuscitate) orders and those who do not.

Data set: **Ch 13 – Exercise 07A.sav**

Codebook

Variable:	DNR
Definition:	Outcome variable: Patient's DNR status
Type:	Categorical (0 = Not DNR, 1 = DNR)
Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)
Variable:	Race
Definition:	Predictor variable: Race
Type:	Categorical (0 = African American, 1 = Asian, 2 = Caucasian, 3 = Latino, 4 = Other)
Variable:	Disease
Definition:	Predictor variable: Primary diagnosis
Type:	Categorical (0 = AIDS, 1 = Cancer, 2 = Cardiac, 3 = Dementia, 4 = Pulmonary, 5 = Stroke, 6 = Other)
Variable:	Religion
Definition:	Predictor variable: Religion
Type:	Categorical (0 = Atheist, 1 = Buddhist, 2 = Catholic, 3 = Hindu, 4 = Jewish, 5 = Other)

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [p value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 07B.sav.**

Exercise 13.8

A nurse in the Infection Control Department wants to investigate the characteristics of those who develop nosocomial infections.

Data set: **Ch 13 – Exercise 08A.sav**

Codebook

Variable:	Nosocomial_infection
Definition:	Outcome variable: Did the patient develop a nosocomial infection?
Type:	Categorical (0 = No infection, 1 = Infection)

Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

Variable:	Ward
Definition:	Predictor variable: Ward
Type:	Categorical (0 = ICU, 1 = CCU, 2 = 1A, 3 = 1B, 4 = 2A, 5 = 2B)

Variable:	LOS
Definition:	Predictor variable: Length of stay in hospital (in days)
Type:	Continuous

Variable:	Surgery
Definition:	Predictor variable: Did the patient have surgery?
Type:	Categorical (0 = Surgery, 1 = No surgery)

   a. Write the hypotheses.

   b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

   c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

   d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 08B.sav.**

Exercise 13.9

The Patient Safety Board has recruited you to determine the factors associated with patient falls during hospitalization.

Data set: **Ch 13 – Exercise 09A.sav**

Codebook

Variable:	Fall
Definition:	Outcome variable: Did the patient fall?
Type:	Categorical (0 = No fall, 1 = Fell)

Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

Variable:	Ward
Definition:	Predictor variable: Ward
Type:	Categorical (0 = 1A, 1 = 1B, 2 = 2A, 3 = 2B)

Variable:	LOS
Definition:	Predictor variable: Length of stay in hospital (in days)
Type:	Continuous

Variable:	Surgery
Definition:	Predictor variable: Did the patient have surgery?
Type:	Categorical (0 = Surgery, 1 = No surgery)

   a. Write the hypotheses.

   b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 09B.sav.**

Exercise 13.10

The Transplant Committee wants to gain a better understanding of those who opt to be an organ donor.

Data set: **Ch 13 – Exercise 10A.sav**

Codebook

Variable:	Organ_donor
Definition:	Outcome variable: Is the person an organ donor?
Type:	Categorical (0 = Not organ donor, 1 = Organ donor)

Variable:	Gender
Definition:	Predictor variable: Gender
Type:	Categorical (0 = Female, 1 = Male)

Variable:	Age
Definition:	Predictor variable: Age
Type:	Continuous

Variable:	Religion
Definition:	Predictor variable: Religion
Type:	Categorical (0 = Atheist, 1 = Buddhist, 2 = Catholic, 3 = Hindu, 4 = Jewish, 5 = Other)

Variable:	SES
Definition:	Predictor variable: Socioeconomic status
Type:	Categorical (0 = Lower class, 1 = Middle class, 2 = Upper class)

a. Write the hypotheses.

b. Run each criterion of the pretest checklist (sample size, normality, multicollinearity) and discuss your findings.

c. Run the logistic regression analysis and document your findings (odds ratios and Sig. [*p* value], hypotheses resolution).

d. Write an abstract under 200 words detailing a summary of the study, the logistic regression analysis results, hypothesis resolution, and implications of your findings.

Repeat this exercise using data set: **Ch 13 – Exercise 10B.sav.**