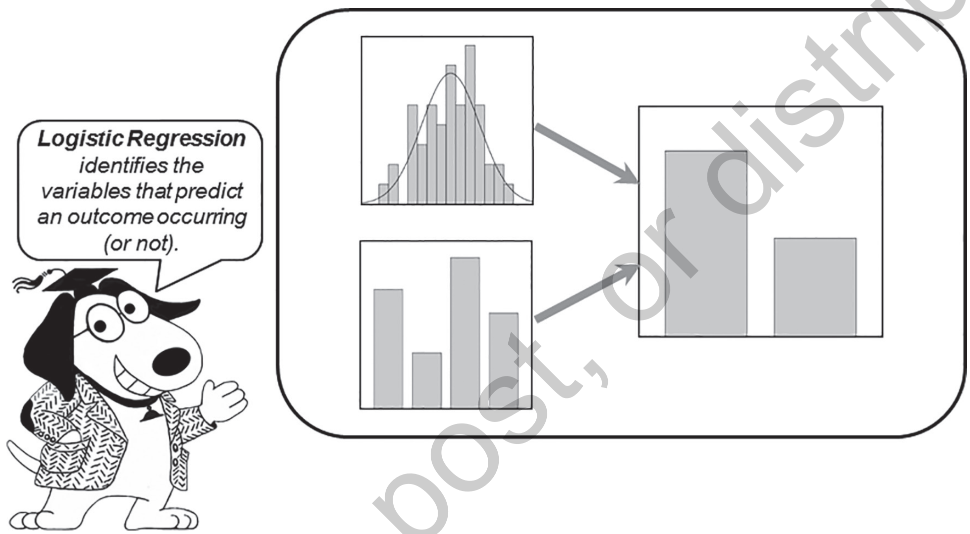
# CHAPTER 13

# Logistic Regression




*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 do the following:

- 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.
- Order a logistic regression test.
- Comprehend the logistic regression $R^2$ 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.
- Write an appropriate abstract.

## WHEN TO USE THIS STATISTIC

### Guidelines for Selecting the Logistic Regression Test

**Overview:** This statistic indicates which variables predict a dichotomous (two-category) outcome.

**Variables:** This statistic can accommodate multiple continuous and categorical predictor variables with one dichotomous outcome variable for each record.

**Results:** After a group of smokers engaged in a smoking cessation program, we gathered data detailing their baseline daily smoking rates, age, gender, race, and income. Logistic regression analysis revealed 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 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). For every additional cigarette smoked per day at baseline, the odds of quitting smoking decreased by 5.4% ($p = .015$; 95% CI .90, .99, $\alpha = .05$).

## 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 the following: did/did not pass an exam, did/did not get a job, did/did not achieve a goal, did/did not submit an assignment on time, and does/does not own a pet.

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 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: Smoking_status

Definition: [Outcome] Smoking status at conclusion of smoking cessation intervention

Type: Categorical

0 = Still smoking

**1 = Quit smoking [←*BASIS FOR MODEL*]**

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

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.

NOTE: If any of the pretest checklist criteria are not satisfied, proceed with the analysis, but concisely discuss such anomalies in the Results or Limitations sections of the documentation so the reader can more plausibly interpret the precision of the findings.

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* (NOTE: This is the same procedure that is used to determine the minimum *n* for multiple regression):

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).
6. Move the outcome variable (*Smoking_Status*) into the *Variable(s)* window (Figure 13.3).
7. Click on *OK*.

**Figure 13.2** To determine the *n* of the data set, click on *Analyze, Descriptive Statistics, Frequencies.*

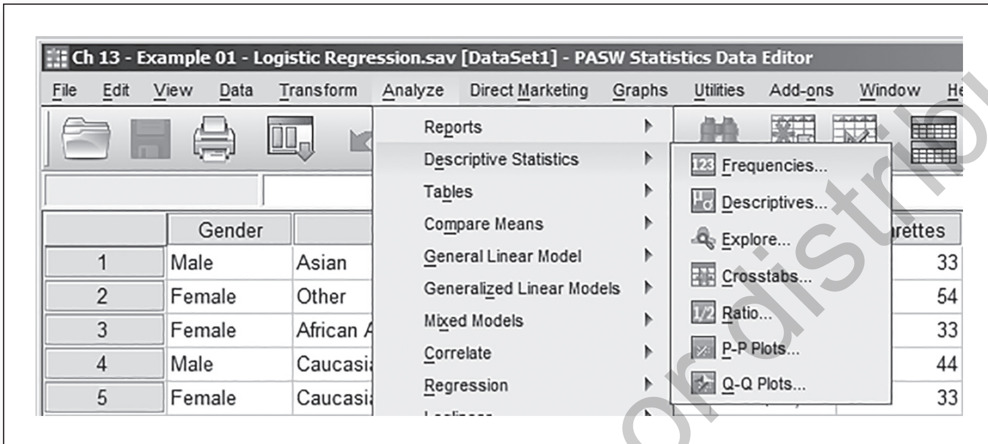
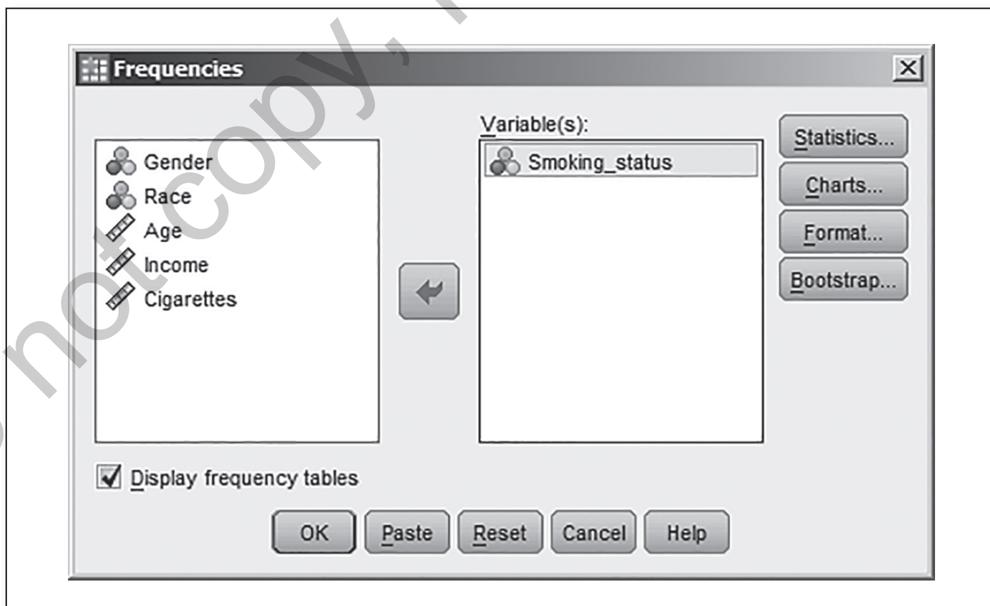



**Figure 13.3** On the *Frequencies* menu, move the outcome variable (*Smoking_status*) into the *Variable(s)* window.



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 | |

**Figure 13.4** To order histograms with normal curves for the continuous variables, click on *Analyze, Descriptive Statistics, Frequencies*.

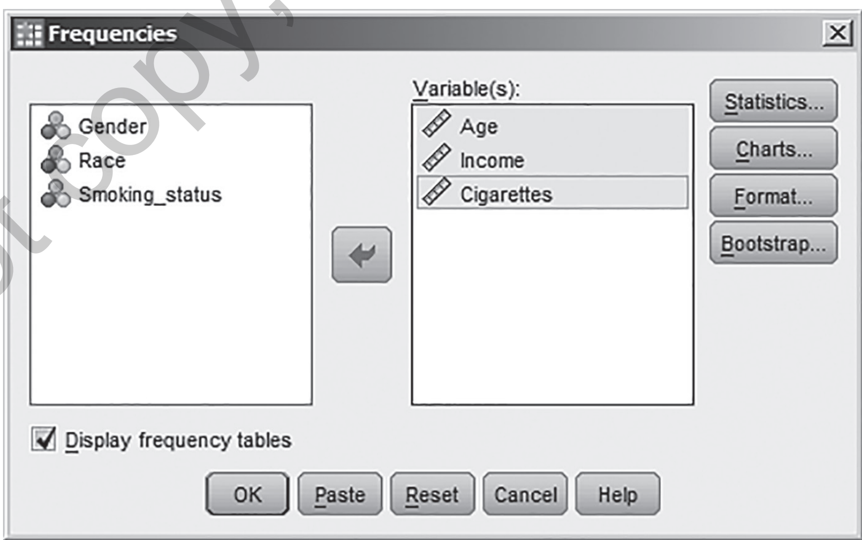


**Figure 13.5** Histogram for *Age*.

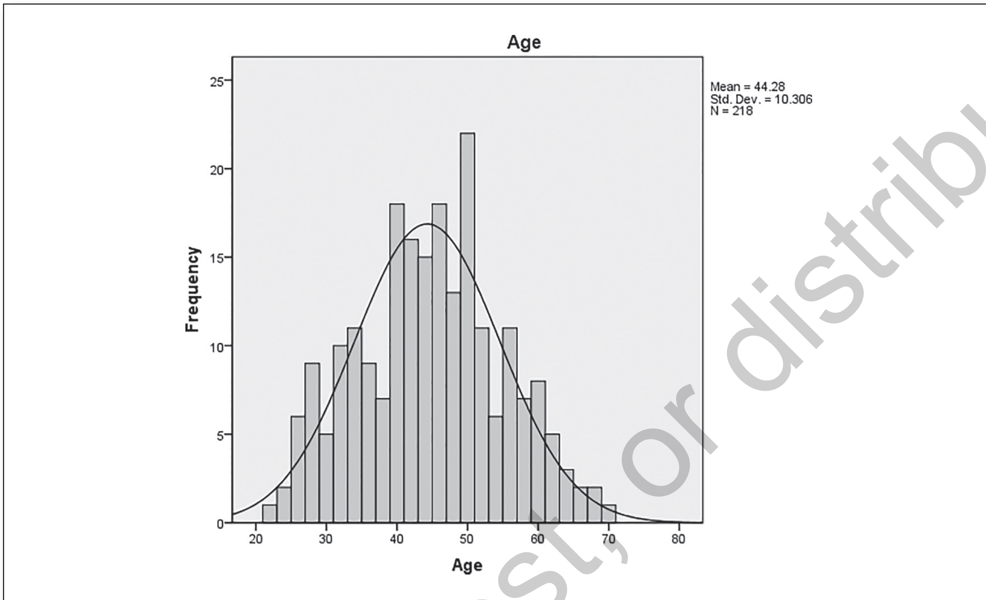



**Figure 13.6** Histogram for *Income*.

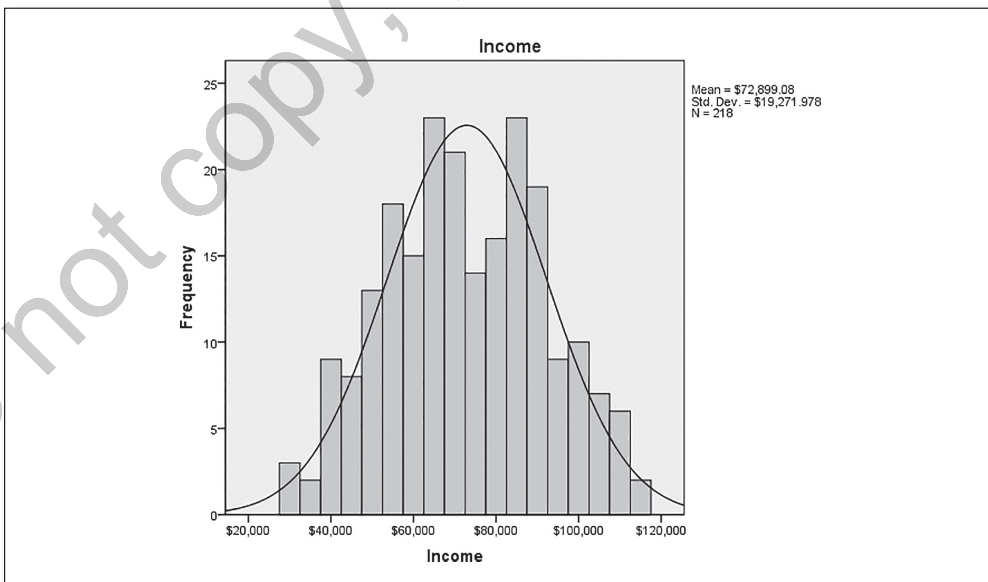
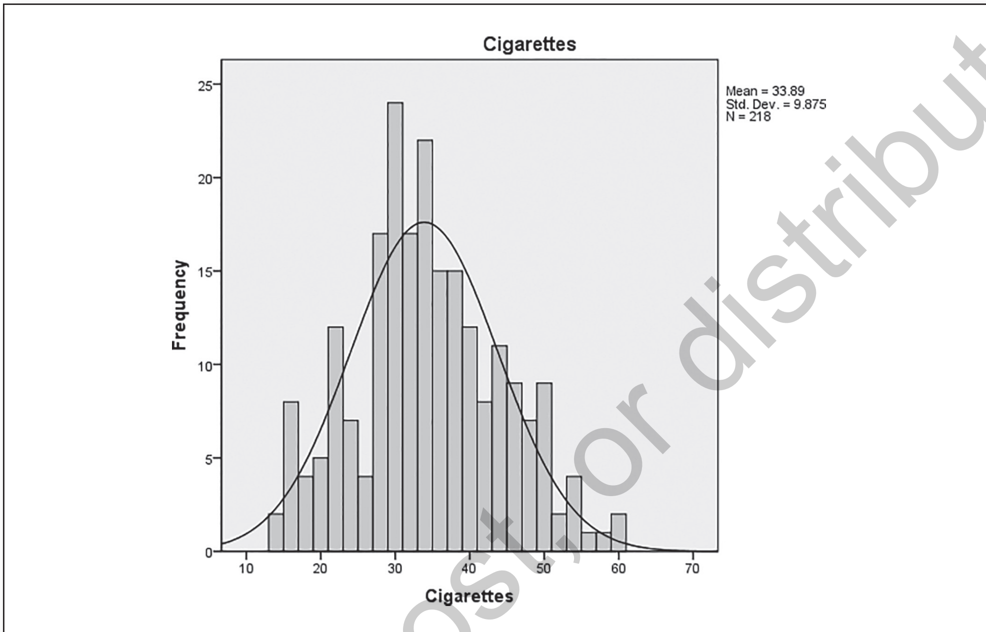


**Figure 13.7** Histogram for *Cigarettes*.






## 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 52; at the ★ icon; move *Age, Income, Cigarettes* into the *Variable(s)* window (Figure 13.4).

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.

## Pretest Checklist Criterion 3—Multicollinearity

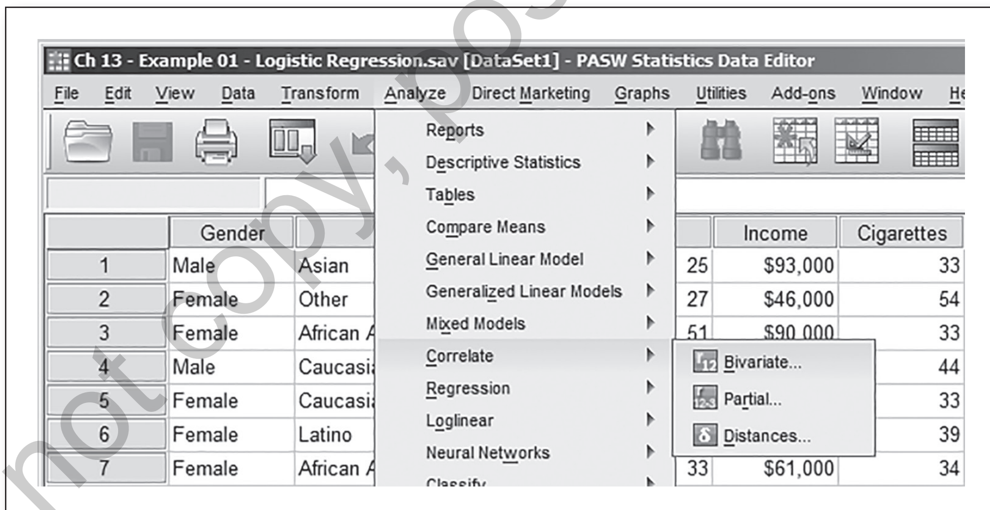
As discussed in Chapter 12, *multicollinearity* describes continuous variables that are very highly correlated. Loading two such variables into a logistic regression model essentially constitutes double-loading the processor; *checking that we do not have multicollinearity* assures us that each continuous variable that we intend to load into the logistic regression model is (statistically) unique. As a rule, in logistic regression, 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).
2. On the *Bivariate Correlations* menu (Figure 13.9), move the continuous variables (*Age, Income, Cigarettes*) into the *Variables* window.
3. Click on *OK*.

**Figure 13.8** Check for multicollinearity; click on *Analyze, Correlate, Bivariate*.




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.

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

**Figure 13.9** *Bivariate Correlation* menu—load *Age, Income,* and *Cigarettes* into *Variables* window.

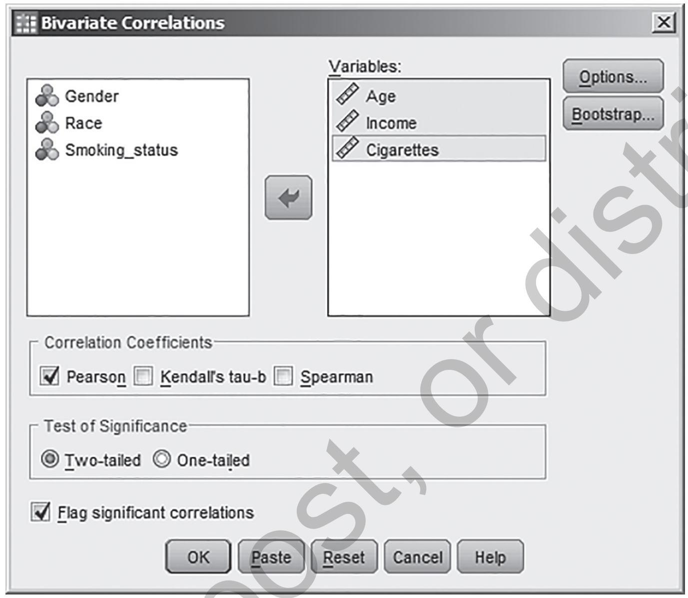



**Table 13.3** *Correlations* table shows the Pearson correlations for each pair of continuous variables (*Age* : *Income*, *Age* : *Cigarettes*, *Income* : *Cigarettes*).

**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 *r* |
|---|---|
| Age : Income | .073 |
| Age : Cigarettes | −.250 |
| Income : Cigarettes | −.095 |

## Test Run

1. On the main SPSS menu, click on *Analyze, Regression, Binary Logistic* (Figure 13.10).
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 13.10** To run a logistic regression, click on *Analyze, Regression, Binary Logistic.*

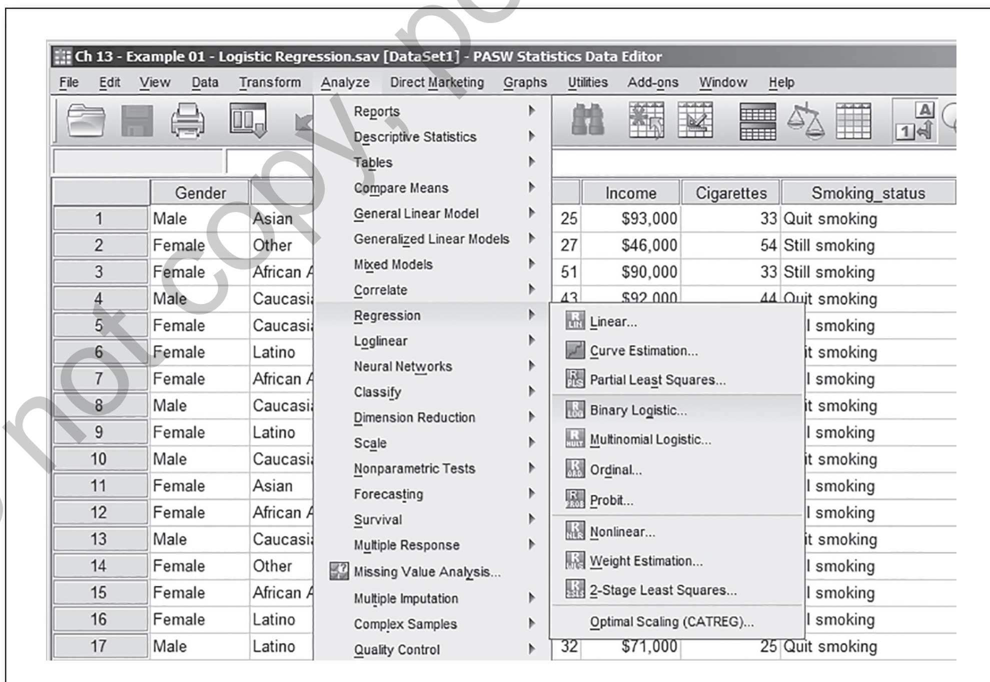
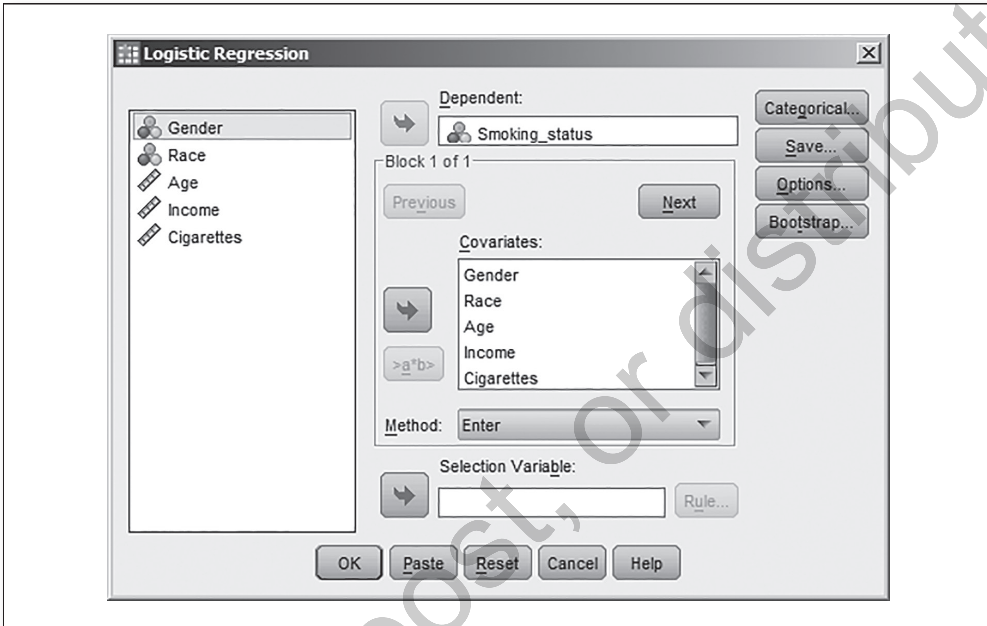


**Figure 13.11** *Logistic Regression* menu—move outcome variable into the *Dependent* box and predictor variables into the *Covariates* window.




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).
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).
11. Click on *Continue*—this will return you to the *Logistic Regression* menu.
12. Click on *OK.*

**Figure 13.12** *Logistic Regression: Define Categorical Variables* menu—move categorical predictor variables into the *Categorical Variables* window.

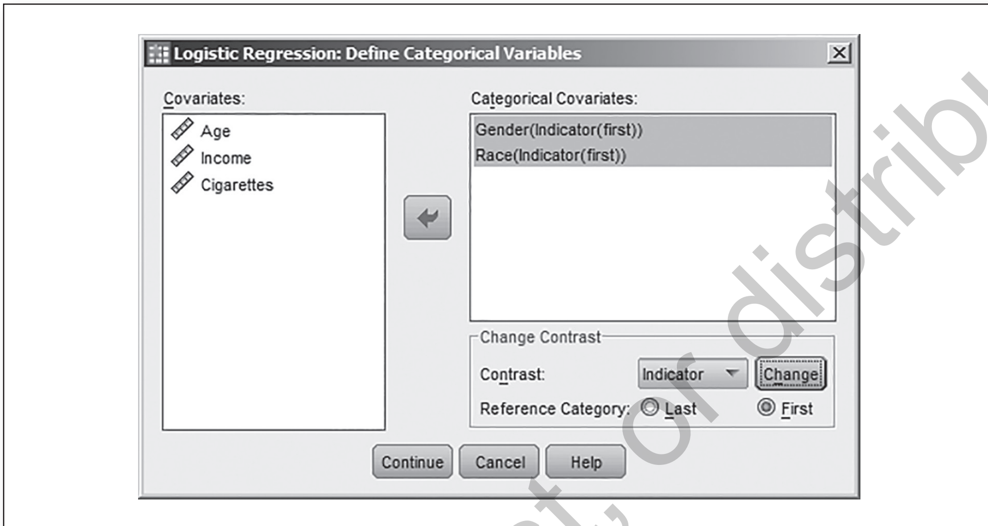
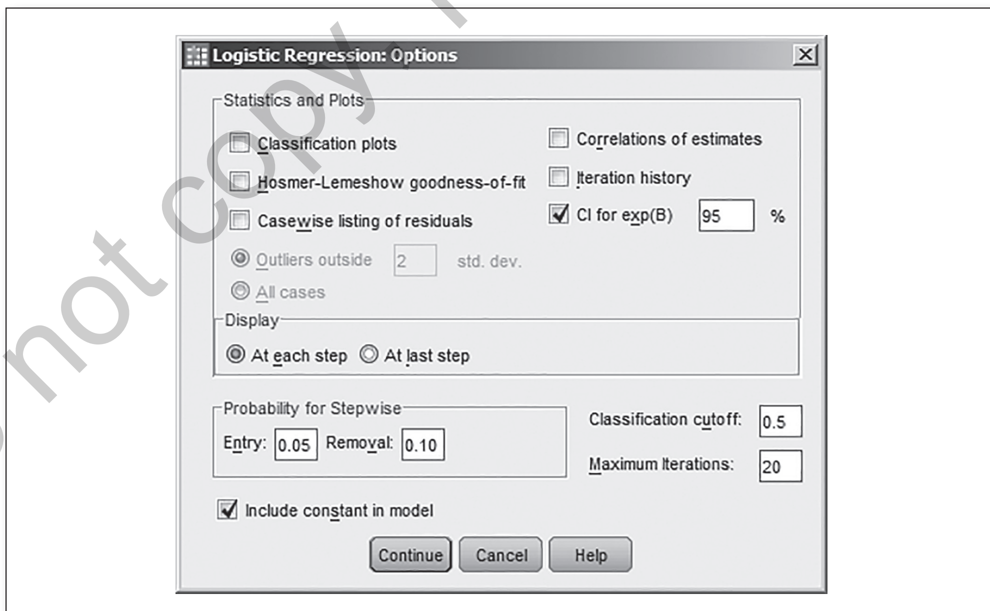



**Figure 13.13** *Logistic Regression Options* menu—select the *CI for exp(B)* (confidence interval), using the default value of 95%.



## 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 Logistic Regression $R^2$

Prior to investigating the *Variables in the Equation* table, we will take a short diversion to discuss the $R^2$ statistic in the logistic regression context. In **Chapter 12: Multiple Regression**, we saw that the $R^2$ indicates the extent to which each predictor variable, and the overall set of predictors, correlate with the (continuous) outcome variable. 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 13.6** *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.

**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 351), 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.

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.

**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.

## $H_0$ Hypothesis Resolution

The *Omnibus Tests of Model Coefficients* table indicates a Sig. (*p*) value of .000. Because this is less than the .05 $\alpha$ 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.

## Documentation Overview

### Abstract

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, because 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 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 $p \le .05$. Alternatively, you may wish to provide detailed discussion of only those categories wherein $p \le .05$ (*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.

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:

**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 | 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 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; because 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 decrease 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; because 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 increase 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; this is greater than the .05 $\alpha$ level, so *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 $p \le .05$. Alternatively, you may wish to provide detailed discussion of only those categories wherein $p \le .05$ 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).

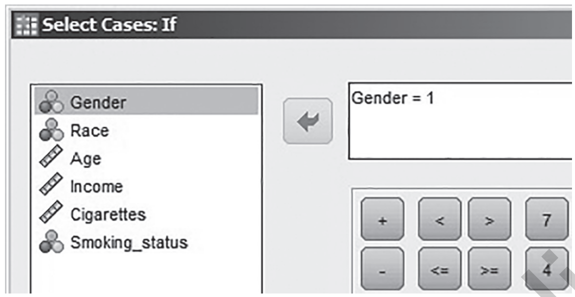
## Model 2: Selective Results

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*.

**Figure 13.14** Selecting cases where *Gender* = 1 (Male).




4. On the *Select Cases: If* menu, enter *Gender* = *1* (Figure 13.14).
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 1[a] | 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 because (now) *Gender* contains only one (selected) value: 1, signifying *Males* (only). Because 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 because 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 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 13.12). 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 the 0s were replaced with 4s, all the 4s were replaced with 0s, and all 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*.

**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_O |
|---|---|
| 0 = African American **[← *REFERENCE*]** | 0 = Other **[← *REFERENCE*]** |
| 1 = Asian | 1 = Asian |
| 2 = Caucasian | 2 = Caucasian |
| 3 = Latino | 3 = Latino |
| 4 = Other | 4 = African American |

The step-by-step instructions for this recoding procedure are detailed in **Chapter 14: Supplemental SPSS Operations**, on page 411 at the ★ icon.

Notice that the figures for *Race* are different from the initial results, because 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:

**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_0, Age, Income, Cigarettes.

### 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, these odds ratios are less than 1, so 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).*

## 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
- Logistic regression $R^2$ 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: Did this person get sick with the flu this season?

Type: Categorical

0 = Got the flu

1 = No flu **[←*BASIS FOR MODEL*]**

Variable: Gender

Definition: [Predictor] Gender

Type: Categorical

0 = Female **[←*REFERENCE*]**

1 = Male

Variable: Flu_shot

Definition: [Predictor] Did person have a flu shot this season?

Type: Categorical

0 = Got a flu shot **[←*REFERENCE*]**

1 = Did not get a flu shot

Variable: Chronic_disease

Definition: [Predictor] Does the person have chronic disease(s)?

Type: Categorical

0 = Has chronic disease(s) **[←*REFERENCE*]**

1 = No chronic disease(s)

Variable: Age

Definition: [Predictor] 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] Age

Type: Categorical

0 = Pediatric **[←*REFERENCE*]**

1 = Adult

**Exercise 13.2**

Acme Solar Systems wants to discover the characteristics of those who intend to install solar energy systems in their homes.

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

Codebook

Variable: Install

Definition: [Outcome] Does this customer intend to install a solar energy system within the next 12 months?

Type: Categorical

0 = Will not install solar energy

1 = Will install solar energy **[←*BASIS FOR MODEL*]**

Variable: Age

Definition: [Predictor] Age

Type: Continuous

Variable: Gender

Definition: [Predictor] Gender

Type: Categorical

0 = Female **[←*REFERENCE*]**

1 = Male

Variable: Income

Definition: Annual household income

Type: Continuous

Variable: Neighborhood

Definition: [Predictor] Type of neighborhood

Type: Categorical

0 = Urban **[←*REFERENCE*]**

1 = Rural

Variable: Family

Definition: Number of people living in the household

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 public opinion consultant is interested in the demographics of those who are in favor of capital punishment (death penalty).

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

Codebook

Variable: Death_penalty

Definition: [Outcome] Are you in favor of the death penalty?

Type: Categorical

0 = Anti-death penalty

1 = Pro-death penalty **[←*BASIS FOR MODEL*]**

Variable: Age

Definition: [Predictor] Age

Type: Continuous

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: Religion
Definition: [Predictor] Religion
Type: Categorical
0 = Atheist **[←*REFERENCE*]**
1 = Buddhist
2 = Catholic
3 = Hindu
4 = Jewish
5 = Other

Variable: Education
Definition: [Predictor] Years of education
Type: Continuous (High school = 12, Associate's = 14, Bachelor's = 16, Master's = 18, Doctorate > 18)

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**.

NOTE: The B data set is the same as the A data set with the following modifications:

- The Race variable has been recoded so that Other is now the reference category:

Categorical

0 = Other **[←*REFERENCE*]**

1 = Asian

2 = Caucasian

3 = Latino

4 = African American

- The Religion variable has been recoded so that Other is now the reference category:

Categorical

0 = Other **[←*REFERENCE*]**

1 = Buddhist

2 = Catholic

3 = Hindu

4 = Jewish

5 = Atheist

**Exercise 13.4**

Acme Employment Services wants to evaluate the effectiveness of its "Get That Job" seminars, which consists of experts facilitating sessions designed to enhance resume writing, job search strategies, and interviewing techniques. After 90 days, participants are surveyed to assess their characteristics and outcomes.

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

Codebook

| | |
|---|---|
| Variable: | Employment_status |
| Definition: | [Outcome] Are you currently employed? |

Type: Categorical

0 = Unemployed

1 = Employed **[←*BASIS FOR MODEL*]**

Variable: Age

Definition: [Predictor] Age

Type: Continuous

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: Experience

Definition: [Predictor] Years of experience working in their current field

Type: Continuous

Variable: Applications

Definition: [Predictor] Total number of job applications submitted

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**

A therapist at the Acme College Counseling Center noted a high prevalence of adjustment disorder among incoming freshmen, with depression being the predominate symptom. The clinicians want to determine the characteristics of those most amenable to therapy over a course of 10 sessions.

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

Codebook

Variable: Treatment_effectiveness

Definition: [Outcome] Did the treatment resolve the adjustment disorder?

Type: Categorical

0 = Treatment ineffective

1 = Treatment effective **[←*BASIS FOR MODEL*]**

Variable: Gender

Definition: [Predictor] Gender

Type: Categorical

0 = Female **[←*REFERENCE*]**

1 = Male

Variable: Age

Definition: [Predictor] Age

Type: Continuous

Variable: Units

Definition: [Predictor] Number of units the student is enrolled in

Type: Continuous

Variable: Work

Definition: [Predictor] Number of hours of (nonacademic) work per week

Type: Continuous

Variable: Treatment_modality

Definition: [Predictor] Form of treatment

Type: Categorical

0 = Individual **[←*REFERENCE*]**

1 = Group

Variable: Home

Definition: [Predictor] Living conditions at home

Type: Categorical

0 = Lives with family **[←*REFERENCE*]**

1 = Lives with roommate(s)

2 = Lives alone

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 technology firm wants to determine the characteristics of potential customers for a new voice-activated home entertainment system.

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

Codebook

Variable: Purchase

Definition: [Outcome] Will the person buy this within 6 months?

Type: Categorical
0 = Will not buy it
1 = Will buy it **[←*BASIS FOR MODEL*]**

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: Partner
Definition: [Predictor] Relational status
Type: Categorical
0 = Single **[←*REFERENCE*]**
1 = Partner

Variable: Age
Definition: [Predictor] Age
Type: Continuous

Variable: Income
Definition: [Predictor] Annual income
Type: Continuous

Variable: Brand_ownership
Definition: [Predictor] Does the person already own any other product(s) of this brand

Type: Categorical

0 = Does not own this brand **[←*REFERENCE*]**

1 = Owns this brand

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**.

### Exercise 13.7

Acme Coffee, which currently sells gourmet coffee blends, is now considering selling a single-serve coffee maker that brews a cup of coffee in 30 seconds. They conduct a survey to help identify the characteristics of potential customers for this high-tech coffee brewer.

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

Codebook

Variable: Buy

Definition: [Outcome] Would you consider buying this coffee brewer?

Type: Categorical

0 = No

1 = Yes **[←*BASIS FOR MODEL*]**

Variable: Age

Definition: [Predictor] Age

Type: Continuous

Variable: Gender

Definition: [Predictor] Gender

Type: Categorical

0 = Female **[←*REFERENCE*]**

1 = Male

Variable: Acme_Coffee

Definition: [Predictor] Does the person currently drink Acme Coffee?

Type: Categorical

0 = Doesn't drink Acme Coffee **[←*REFERENCE*]**

1 = Drinks Acme Coffee

Variable: Income

Definition: [Predictor] Annual household income

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**

In an effort to identify the characteristics of incoming high school students who are most vulnerable to dropping out, the research staff gathered data on the senior students at the end of the school year. Based on this data, freshmen who are identified as vulnerable to dropping out will be offered access to free comprehensive tutorial services.

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

Codebook

Variable: HS_completion

Definition: [Outcome] Did the student drop out or graduate?

Type: Categorical

0 = Graduated

1 = Drop-out **[←*BASIS FOR MODEL*]**

Variable: Gender

Definition: [Predictor] Gender

| | |
|---|---|
| Type: | Categorical |
| | 0 = Female **[←*REFERENCE*]** |
| | 1 = Male |
| Variable: | Adjusted_income |
| Definition: | [Predictor] Annual household income ÷ number of people in household |
| Type: | Continuous |
| Variable: | Education_parents |
| Definition: | [Predictor] Highest years of parent's education |
| Type: | Continuous (e.g., High school = 12, Associate's = 14, Bachelor's = 16, Master's = 18, Doctorate > 18) |
| Variable: | Language_skill |
| Definition: | [Predictor] Pre–high school reading and writing skills placement exam |
| Type: | Continuous (1 . . . 30) |
| Variable: | Math_skill |
| Definition: | [Predictor] Pre–high school math skills placement exam |
| Type: | Continuous (1 . . . 30) |

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

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

Codebook

| | |
|---|---|
| Variable: | Organ_donor |
| Definition: | [Outcome] Is the person an organ donor? |
| Type: | Categorical |
| | 0 = Not organ donor |
| | 1 = Organ donor **[←*BASIS FOR MODEL*]** |
| Variable: | Gender |
| Definition: | [Predictor] Gender |
| Type: | Categorical |
| | 0 = Female **[←*REFERENCE*]** |
| | 1 = Male |
| Variable: | Age |
| Definition: | [Predictor] Age |
| Type: | Continuous |
| Variable: | Religion |
| Definition: | [Predictor] Religion |
| Type: | Categorical |
| | 0 = Atheist **[←*REFERENCE*]** |
| | 1 = Buddhist |
| | 2 = Catholic |
| | 3 = Hindu |
| | 4 = Jewish |
| | 5 = Other |
| Variable: | SES |
| Definition: | [Predictor] Socioeconomic status |
| Type: | Categorical |
| | 0 = Lower class **[←*REFERENCE*]** |
| | 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 09B.sav**.

**Exercise 13.10**

The Acme Industries Safety Supervisor wants to determine the factors that predict employees passing the annual required site safety competency training course.

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

Codebook

| | |
|---|---|
| Variable: | Test_result |
| Definition: | [Outcome] Did the employee pass the annual safety exam? |
| Type: | Categorical |
| | 0 = Fail |
| | 1 = Pass **[←*BASIS FOR MODEL*]** |
| Variable: | Training_type |
| Definition: | [Predictor] Training type |
| Type: | Categorical |
| | 0 = Workbook **[←*REFERENCE*]** |
| | 1 = Online course |
| | 2 = Simulation lab |
| Variable: | Years |
| Definition: | [Predictor] Years of professional experience |
| Type: | Continuous |
| Variable: | Employment_hours |
| Definition: | [Predictor] Part-time or full-time |

Type: Categorical

0 = Part-time [←***REFERENCE***]

1 = Full-time

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**.