*The text below lists the Stata, SPSS, and R commands needed to reproduce the tables and figures in this volume.*

**\*\*\*\*\*\*\*\*\*\*\*\*\*\* LOGISTIC REGRESSION BOOK ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STATA \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\* NATIONAL HEALTH INTERVIEW SURVEY DATA ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

clear

use "NHIS 2017 Logistic Regression Primer.dta"

set scheme s2mono

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Regression and save predicted outcome

reg Smoker Education Gender

predict pSmoker if e(sample)

\*\* Split on dependent variable

estat sum

\*\* Min and max predicted values

sum pSmoker

\*\* Cases with lowest and highest predicted values

sum pSmoker if Education==18 & Gender==0

sum pSmoker if Education==0 & Gender==1

\*\*Add age and check predicted values

reg Smoker Education Gender i.Agecat

predict p2Smoker if e(sample)

sum p2Smoker

sum p2Smoker Education Gender Age if p2Smoker <-.05

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 2 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*Logistic regression (Table 2.1)

logit Smoker Education Age i.Gender i.Race i.Ethnicity

\*\* Predicted values (Table 2.4)

predict Logit\_Smoker, xb

predict Prob\_Smoker

reg Smoker Education Age i.Gender i.Race i.Ethnicity

predict Reg\_Smoker

sum Smoker Logit\_Smoke Prob\_Smoke Reg\_Smoke if e(sample)

\*\* Margins

logit Smoker Education Age i.Gender i.Race i.Ethnicity

\*\* Marginal effects at means (Table 2.5)

margins , dydx(\*) atmeans

\*\* Marginal effects at representative values (Table 2.6)

margins , dydx(\*) at(Education=12 Age=45 Gender=1 Race=1 Ethnicity=0)

\*\* Mean predicted probabilities for representative value and means

margins , at(Education=12 Age=45 Gender=1 Race=1 Ethnicity=0)

margins , atmeans

\*\* Average marginal effects (Table 2.7)

margins , dydx(\*)

\*\* Gender difference for marginal effects at means

margins Gender, atmeans

\*\* Marginal effects at selected values for education

\*\* Display values of education at varied standard deviations

sum Education if e(sample)

dis r(mean) - 2\*(r(sd))

dis r(mean) - (r(sd))

dis r(mean)

dis r(mean) + (r(sd))

dis r(mean) + 2\*(r(sd))

\*\* Compute marginal effects at standard deviations, min, and max

logit Smoker Education Age i.Gender i.Race i.Ethnicity

margins , dydx(Education) at(Education=(8.215 11.062 13.909 16.756 19.603))

margins , dydx(Education) at(Education=(0 13.909 18))

\*\* Graphs for Marginal Effects

\*\* Non-Linearity: Average marginal effect of education (Figure 2.2)

margins , dydx(Education) at(Education=(0(1)18))

marginsplot

graph export figure\_2\_2.png

\*\* Non-additivity: Average marginal effect of gender by age (Figure 2.3)

margins , dydx(Gender) at(Age=(26(1)85))

marginsplot

graph export figure\_2\_3.png

\*\* Average marginal effect of education by age (Figure 2.4)

margins , dydx(Education) at(Age=(26(1)85))

marginsplot

graph export figure\_2\_4.png

\*\* Graphs for Predicted Probabilities

\*\* Predicted probabilities for gender by age (Figure 2.5)

margins Gender, at(Age=(25(5)85))

marginsplot

graph export figure\_2\_5.png

\*\* Predicted probabilities for ethnicity by education (Figure 2.6)

margins Ethnicity, at(Education=(0(1)18))

marginsplot

graph export figure\_2\_6.png

\*\* Predicted probabilities with regression (Figure 2.7)

reg Smoker Education Age i.Gender i.Race i.Ethnicity

margins Ethnicity, at(Education=(0(1)18))

marginsplot

graph export figure\_2\_7.png

\*\* Standardization

\*\* Semi-Standardized by standardizing independent variables

\*\* Create dummy variables for race

tab Race, gen(Race)

sum Race\*

\*\* Standardize each predictor

quietly logit Smoker Education Age i.Gender i.Race i.Ethnicity

foreach var of varlist Education Age Gender Race2-Race5 Ethnicity {

egen z`var' = std(`var') if e(sample)

}

rename zRace2 zAfricanAmer

rename zRace3 zNativeAmer

rename zRace4 zAsianAmer

rename zRace5 zMultiRace

sum z\*

\*\* Logistic regression with standardized predictors (Table 2.8)

logit Smoker zEducation zAge zGender zAfricanAmer-zMultiRace zEthnicity

\*\* Calculate standard deviation of latent continuous outcome

\*\* Variance of predicted logit value

logit Smoker Education Age i.Gender i.Race i.Ethnicity

sum Logit\_Smoker if e(sample)

dis r(sd)^2

\*\* Variance of logit distribution

dis ((\_pi^2)/3)

\*\* Standard deviation of continuous latent outcome

sca sdy = sqrt((r(sd)^2) + ((\_pi^2)/3))

dis sdy

\*\* Fully standardized coefficient for education

sum Education if e(sample)

sca sdx = r(sd)

dis sdx

dis (\_b[Education])\*(sdx/sdy)

\*\* Fully standardized coeficients from SPOST (Table 2.9)

listcoef , std

\*\* Interaction runs (Table 2.10 and Figure 2.8)

logit Smoker c.Education##Ethnicity Age i.Gender i.Race

margins , dydx(Ethnicity) at(Education=(0(1)18))

marginsplot

graph export figure\_2\_8.png

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 4 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Probit model (Table 4.1 and 4.3)

probit Smoker Education Age i.Gender i.Race i.Ethnicity

listcoef, std

\*\* Logit comparison (Table 4.2 and 4.3)

logit Smoker Education Age i.Gender i.Race i.Ethnicity

listcoef, std

\*\* Probit average marginal effects (Table 4.4)

quietly probit Smoker Education Age i.Gender i.Race i.Ethnicity

margins , dydx(\*)

exit, clear

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STATA \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\* GENERAL SOCIAL SURVEY DATA ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

clear

use "GSS 7216 Logistic Regression Primer.dta"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Regression

reg Grass\_Legal Education i.Gender Time

estat sum

\*\* Odds for males and females

tab Grass\_Legal Gender if year==2016, col

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Basic model (Table 3.6)

logit Grass\_Legal Education Time Age i.Gender i.SizeOfPlace

dis ( -22045.785 - -20193.291)

dis ( -22045.785 - -20193.291) \* -2

\*\* Reduced model (Table 3.8)

logit Grass\_Legal Education Time Age i.Gender

dis 3704.99-3518.51

dis chi2(186.48, 5)

\*\* Fitstat statistics (Table 3.10)

logit Grass\_Legal Education Time Age i.Gender i.SizeOfPlace

fitstat

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 5 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*Support for environmental spending

\*\* Check for zero cells

tab Spend\_Env Gender

tab Spend\_Env SizeOfPlace

\*\* Logged odds (Table 5.1)

ologit Spend\_Env Education Time Age i.Gender i.SizeOfPlace

\*\* Odds ratios (Table 5.2)

ologit Spend\_Env Education Time Age i.Gender i.SizeOfPlace, or

\*\* Predicted probabilities (Table 5.3)

quietly ologit Spend\_Env Education Time Age i.Gender i.SizeOfPlace

predict pSpend\_Env1 pSpend\_Env2 pSpend\_Env3

tab Spend\_Env if e(sample)

sum pSpend\_Env1-pSpend\_Env3 if e(sample)

\*\* Note on predicted probabilities by observed outcome

tabstat pSpend\_Env1-pSpend\_Env3, by(Spend\_Env) stat(mean sd), if e(sample)

\*\* Average marginal effects (Table 5.4)

margins , dydx(\*)

\*\* Test proportional odds or parallel regression assumption (Table 5.5)

brant, detail

exit, clear

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STATA \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* WORLD VALUES SURVEY DATA ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

clear

use "WVS Wave6 Logistic Regression Primer.dta"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 5 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\* Select India

keep if V2 == 356

\*\* Five-category outcome measure

tab Most\_Serious\_Problem

\*\* Multinomial logistic regression

\*\* Base model (Table 5.8)

mlogit Most\_Serious\_Problem Educ\_Degree Age i.Gender Size\_Town

\*\* Relative risk ratios (Table 5.9)

mlogit Most\_Serious\_Problem Educ\_Degree Age i.Gender Size\_Town, rrr

\*\* Change base category to pollution (Table 5.10)

mlogit Most\_Serious\_Problem Educ\_Degree Age i.Gender Size\_Town, base(5) rrr

\*\* Average marginal effect for pollution as base (Table 5.11)

margins , dydx(\*)

exit, clear

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* SPSS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\* NATIONAL HEALTH INTERVIEW SURVEY DATA ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\* .

\*\* Use NHIS data from Stata.

GET STATA file="NHIS 2017 Logistic Regression Primer.dta".

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 2 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*.

\*\* Logistic regression model (Table 2.2).

LOGISTIC REGRESSION VARIABLES Smoker

/METHOD=ENTER Education Age Gender Race Ethnicity

/CONTRAST (Gender)=Indicator(1)

/CONTRAST (Race)=Indicator(1)

/CONTRAST (Ethnicity)=Indicator(1)

/SAVE=PRED

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* SPSS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\* GENERAL SOCIAL SURVEY DATA ANALYSIS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*.

\*\* Use GSS data from Stata.

GET STATA file="GSS 7216 Logistic Regression Primer.dta".

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*.

\*\* Legalization of marijuana logistic regression (Table 3.9).

LOGISTIC REGRESSION VARIABLES Grass\_Legal

/METHOD=ENTER Education Time Age Gender SizeOfPlace

/CATEGORICAL=Gender SizeOfPlace

/CONTRAST (Gender)=Indicator(1)

/CONTRAST (SizeOfPlace)=Indicator(1)

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

\*\*\*\*\*\*\*\*\*\*\*\*\*\* CHAPTER 5 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*.

\*\* Support for gay marriage ordinal logistic regression (Table 5.6).

PLUM Gay\_Marry BY Gender Marital WITH Education Time Age

/CRITERIA=CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5) PCONVERGE(1.0E-6) SINGULAR(1.0E-8)

/LINK=LOGIT

/PRINT=FIT PARAMETER SUMMARY TPARALLEL.

\*\* Support for envir spending multinomial logistic regression (Table 5.12).

NOMREG Spend\_Env (BASE=LAST ORDER=ASCENDING) BY Gender Marital WITH Education Time Age

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0) PCONVERGE(0.000001) SINGULAR(0.00000001)

/MODEL

/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR) REMOVALMETHOD(LR)

/INTERCEPT=INCLUDE

/PRINT=PARAMETER SUMMARY LRT CPS STEP MFI.

#################### R ##########################

#### NATIONAL HEALTH INTERVIEW SURVEY DATA ANALYSIS ####

library(margins)

library(ggeffects)

library(snakecase)

library(effectsizescr)

#Calling csv file into data frame

nhis <- read.csv("NHIS 2017 Logistic Regression Primer.csv")

#Make factor variables

nhis$Agecat.f <- factor(nhis$Agecat)

nhis$Race.f <- factor(nhis$Race)

nhis$Gender.f <- factor(nhis$Gender)

nhis$Ethnicity.f <- factor(nhis$Ethnicity)

####### Chapter 2 ##############

#Logistic regression (Table 2.3)

mylogit <- glm(Smoker ~ Education + Age + Gender.f +

Race.f + Ethnicity.f, data = nhis, family = "binomial")

summary(mylogit)

exp(cbind(coef(mylogit), confint(mylogit)))

#FINISH

################ R #####################

###### GENERAL SOCIAL SURVEY DATA ANALYSIS ####

library(MASS)

library(nnet)

library(broom)

#calling csv file into data frame

gss <- read.csv("GSS 7216 Logistic Regression Primer.csv")

summary(gss)

sd(na.omit(gss$Age))

#make agecat and raceth a factor

gss$Gender.f <- factor(gss$Gender)

gss$Marital.f <- factor(gss$Marital)

gss$Gay\_Marry.f <- factor(gss$Gay\_Marry)

gss$Spend\_Env.f <- factor(gss$Spend\_Env)

gss$SizeOfPlace.f <- factor(gss$SizeOfPlace)

gss$Grass\_Legal.f <- factor(gss$Grass\_Legal)

table(gss$Gay\_Marry)

table(gss$Spend\_Env)

table(gss$Grass\_Legal)

######## Chapter 3 ##############

#logistic regression legalization of marijuana (Table 3.11)

logit <- glm(Grass\_Legal.f ~ Education + Time + Age + Gender.f +

SizeOfPlace.f, data = gss, family = "binomial")

summary(logit)

####### Chapter 5 ##############

#ordinal logistic regression support for gay marriage (Table 5.7)

ologit <- polr(Gay\_Marry.f ~ Education + Time + Age +

Gender.f + Marital.f , data = gss,

Hess = TRUE)

##show coefficients plus null and model deviance

summary(ologit)

##add pvalues

ologit\_pvalues <- coef(summary(ologit))

pval <- pnorm(abs(ologit\_pvalues[, "t value"]),lower.tail = FALSE)\* 2

ologit\_pvalues <- cbind(ologit\_pvalues, "p value" = round(pval,3))

ologit\_pvalues

##odds ratio

exp(coef(ologit))

ologit\_or <- exp(coef(ologit))

ci <- confint(ologit)

exp(cbind(OR = coef(ologit), ci))

##multinomial logistic regress support for envir spending (Table 5.13)

# gss$Spend\_Env2.f <- relevel(gss$Spend\_Env.f, ref = "2")

# mlogit <- multinom(Spend\_Env2.f ~ Education + Time + Age + Gender.f +

# SizeOfPlace.f, data = gss)

# summary(mlogit)

# broom::tidy(mlogit, exponentiate=F)