Question 1

Joseph Froelicher

11/12/2020

# Part A

Where, is the y-intercept, is smoking status (CURSMOKE), is body mass index (BMI), is prevalent coronary heart disease (PREVCHD), is prevalent myocardial infarction (PREVMI), is prevalent stroke (PREVSTRK), and is prevalent hypertensive (PREVHYP). through are the slopes for each corresponding variable mentioned.

# Part B

fram <- fram[,c(3, 7, 9, 15, 17, 18, 19)]  
dim(fram)

## [1] 4434 7

for(column in 1:dim(fram)[2]) {  
 fram <- fram[!is.na(fram[,column]),]  
}  
  
dim(fram)

## [1] 4364 7

# Part C

fit\_full <- glm(TOTCHOL ~ CURSMOKE + BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP, data = fram)  
fit\_empty <- glm(TOTCHOL ~ 1, data = fram)  
anova(fit\_full, fit\_empty, test = "F")

## Analysis of Deviance Table  
##   
## Model 1: TOTCHOL ~ CURSMOKE + BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP  
## Model 2: TOTCHOL ~ 1  
## Resid. Df Resid. Dev Df Deviance F Pr(>F)   
## 1 4357 8410206   
## 2 4363 8700437 -6 -290231 25.06 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

There is evidence to suggest (p < 0.0001) that we should reject the null hypothesis that all of our coefficients in the full multiple linear regression model have a slope of zero. At least one coefficient is non zero.

# Part D

fit\_nocardio <- glm(TOTCHOL ~ CURSMOKE + BMI, data = fram)  
anova(fit\_full, fit\_nocardio, test = "F")

## Analysis of Deviance Table  
##   
## Model 1: TOTCHOL ~ CURSMOKE + BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP  
## Model 2: TOTCHOL ~ CURSMOKE + BMI  
## Resid. Df Resid. Dev Df Deviance F Pr(>F)   
## 1 4357 8410206   
## 2 4361 8561019 -4 -150814 19.533 6.022e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

There is evidence to suggest that we should reject the null hypothesis (p < 0.0001) that all of our cardiovascular condition slopes are zero. At least one of the cardiovascular condition slopes is non-zero.

# Part E

fit\_nosmoke <- glm(TOTCHOL ~ BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP, data = fram)  
anova(fit\_full, fit\_nosmoke, test = "F")

## Analysis of Deviance Table  
##   
## Model 1: TOTCHOL ~ CURSMOKE + BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP  
## Model 2: TOTCHOL ~ BMI + PREVCHD + PREVMI + PREVSTRK + PREVHYP  
## Resid. Df Resid. Dev Df Deviance F Pr(>F)  
## 1 4357 8410206   
## 2 4358 8414969 -1 -4763.1 2.4676 0.1163

There is not significant evidence to suggest that we should reject the null hypothesis (p = 0.1163) that all of our cardiovascular conditions slopes are zero, so we will fail to reject the null hypothesis. The slope for smoking status is not significantly different than zero.

# Part F

vif(fit\_full)

## CURSMOKE BMI PREVCHD PREVMI PREVSTRK PREVHYP   
## 1.035263 1.123308 1.803381 1.783301 1.012782 1.121380

None of our variance inflation factors for any of the variables are greater than our usual cutoff of 10. In fact they are not even notably close. While based on the description of the variables, we may be concerned with collinearity between similar cardiovascular conditions, the measures of variance inflation factor do not suggest that there is any collinearity between the independent variables.

# Part G

In the presence of smoking status and the four cardiovascular conditions (heart disease, myocardial infarction, stroke, and hypertension), a unit increase in body mass index (BMI) is accounting for a 0.8465525 increase in total cholesterol.