Homework 9\_Joel Parker

1. Please go through the tutorial titled Multivariable Logistic Models UsingSAS/Stata“, which can be downloaded from the d2l folder ” or ". Repeatall steps of the analysis and make sure you understand everything in the document,including all commands, options, outputs, etc. Please feel free to contact the instructorfor any questions. You don’t need to turn in anything for this part of the assignment.

2.) . The Women’s Health Study randomly assigned 39,876 initially healthy women aged 45years or older to receive 100 mg of aspirin on alternate days or placebo and monitoredthem for 10 years for a first major cardiovascular event. The table below shows the results stratified by age at randomization. The data are also in the lewhsasa.dat posted on the course website. The variables include: age (scored 50, 60, or 70); asa(0=placebo, 1=aspirin); cvd (0=no, 1=yes); and count (number of women) \* Use logistic regression to characterize the relationship between aspirin assignmentand the odds of CVD.

1. **Use logistic regression to characterize the relationship between aspirin assignmentand the odds of CVD.**

* Obtain the crude odds ratio estimate and its 95% confidence interval;

crude <- glm(cvd~asp, data = whsa, family = binomial(link = "logit"))  
crude$coefficients

## (Intercept) asp   
## -3.61639115 -0.09205454

exp(crude$coefficients)

## (Intercept) asp   
## 0.02687951 0.91205541

*The crude odds ratio is .912.*

* Evaluate whether age confounds the CVD-aspirin relationship by using dummyvariables for age categories.

confound <- glm(cvd~as.factor(age)+asp, data=whsa, family=binomial(link = "logit"))  
summary(confound)

##   
## Call:  
## glm(formula = cvd ~ as.factor(age) + asp, family = binomial(link = "logit"),   
## data = whsa)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4028 -0.2467 -0.1686 -0.1609 2.9507   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.24692 0.06387 -66.495 <2e-16 \*\*\*  
## as.factor(age)2 0.86331 0.07699 11.214 <2e-16 \*\*\*  
## as.factor(age)3 1.77586 0.08162 21.759 <2e-16 \*\*\*  
## asp -0.09337 0.06461 -1.445 0.148   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9338.9 on 39875 degrees of freedom  
## Residual deviance: 8888.8 on 39872 degrees of freedom  
## AIC: 8896.8  
##   
## Number of Fisher Scoring iterations: 7

*The p-values for the 55-64 and older than 65 age groups are less than .2 in both categories thus they need to be adjusted for.*

* Calculate the age-adjusted odds ratio estimate and confidence interval.

exp(confound$coefficients)

## (Intercept) as.factor(age)2 as.factor(age)3 asp   
## 0.0143083 2.3710062 5.9053607 0.9108603

exp(-.09337 + c(-1.96\*.148, 1.96\*1.48))

## [1] 0.6815062 16.5672855

The adjusted odds ratio is 0.91 with a 95% confidence interval of (.6815, 16.5672).

1. **Evaluate possible effect modification by age.**

* Obtain age-speciffic odds ratio estimates and corresponding 95% confidence intervals for aspirin use (hint: in SAS you could use theby age statement in proc logistic after sorting the dataset byage; in Stata you could useifin thelogitcommand to specify the age group, e.g.,logit cvd asa ifage==50 [fw=count], or)

#### Group 1  
summary(group\_1 <-glm(cvd~asp, data = whsa[whsa$age==1,],   
 family = binomial(link = "logit")))

##   
## Call:  
## glm(formula = cvd ~ asp, family = binomial(link = "logit"), data = whsa[whsa$age ==   
## 1, ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.1653 -0.1653 -0.1643 -0.1643 2.9368   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.29902 0.07934 -54.182 <2e-16 \*\*\*  
## asp 0.01294 0.11187 0.116 0.908   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3434 on 24024 degrees of freedom  
## Residual deviance: 3434 on 24023 degrees of freedom  
## AIC: 3438  
##   
## Number of Fisher Scoring iterations: 7

exp(group\_1$coefficients)

## (Intercept) asp   
## 0.01358191 1.01302057

exp(.01294 +c(-1.96\*.11187,1.96\*.11187))

## [1] 0.8135685 1.2613785

summary(group\_2 <-glm(cvd~asp, data = whsa[whsa$age==2,],   
 family = binomial(link = "logit")))

##   
## Call:  
## glm(formula = cvd ~ asp, family = binomial(link = "logit"), data = whsa[whsa$age ==   
## 2, ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.2536 -0.2536 -0.2515 -0.2515 2.6341   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.42107 0.07451 -45.913 <2e-16 \*\*\*  
## asp -0.01644 0.10579 -0.155 0.877   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3280.6 on 11753 degrees of freedom  
## Residual deviance: 3280.6 on 11752 degrees of freedom  
## AIC: 3284.6  
##   
## Number of Fisher Scoring iterations: 6

exp(group\_2$coefficients)

## (Intercept) asp   
## 0.03267744 0.98369815

exp(-.016 +c(-1.96\*.106,1.96\*.106))

## [1] 0.799507 1.211380

summary(group\_3 <-glm(cvd~asp, data = whsa[whsa$age==3,],   
 family = binomial(link = "logit")))

##   
## Call:  
## glm(formula = cvd ~ asp, family = binomial(link = "logit"), data = whsa[whsa$age ==   
## 3, ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4226 -0.4226 -0.3636 -0.3636 2.3450   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.37104 0.07904 -29.997 < 2e-16 \*\*\*  
## asp -0.31227 0.12001 -2.602 0.00927 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2176.3 on 4096 degrees of freedom  
## Residual deviance: 2169.5 on 4095 degrees of freedom  
## AIC: 2173.5  
##   
## Number of Fisher Scoring iterations: 5

exp(group\_3$coefficients)

## (Intercept) asp   
## 0.09338314 0.73178031

exp(-.312 +c(-1.96\*.12,1.96\*.12))

## [1] 0.5785675 0.9260751

*The OR for the 45-54 age group is 1.01 with a CI of (0.814, 1.261).* *The OR for the 55-64 age group is .984 with a confidence interval of (0.79995, 1.211380).* *The odds ratio for the 65 and over age group is 0.732 with a confidence interval of (0.58, 0.93)*

* Test for effect modiffcation by first including age in the model as a continuous covariate (scores 50, 60 and 70) and adding an interaction of this continuous variable with aspirin assignment;

whsa\_1 <- whsa  
whsa\_1$age[whsa\_1$age==1] = 50  
whsa\_1$age[whsa\_1$age==2] = 60  
whsa\_1$age[whsa\_1$age==3] = 70  
  
effect\_modif <- glm(cvd~age+ asp+ age\*asp, data = whsa\_1,   
 family = binomial(link = "logit"))  
summary(effect\_modif)

##   
## Call:  
## glm(formula = cvd ~ age + asp + age \* asp, family = binomial(link = "logit"),   
## data = whsa\_1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4162 -0.2480 -0.1666 -0.1617 2.9476   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.151751 0.344000 -26.604 <2e-16 \*\*\*  
## age 0.096415 0.005681 16.971 <2e-16 \*\*\*  
## asp 0.862320 0.493338 1.748 0.0805 .   
## age:asp -0.016034 0.008208 -1.953 0.0508 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9338.9 on 39875 degrees of freedom  
## Residual deviance: 8885.2 on 39872 degrees of freedom  
## AIC: 8893.2  
##   
## Number of Fisher Scoring iterations: 7

*With a pvalue of .05 when modeling age as a continuous variable, we would want to adjust for the interaction of age and asp.*

* Test for effect modiffication using age as a categorical covariate (dummy vari-ables) and adding interactions with aspirin assignment (hint: use likelihoodratio test to check whether the interaction terms are significant or not).

effect\_cat <- glm(cvd~as.factor(age)+asp+as.factor(age)\*asp,  
 data= whsa\_1,  
 family = binomial(link = "logit"))  
effect\_cat\_no <- glm(cvd~as.factor(age)+asp, data=whsa\_1, family=binomial(link = "logit"))  
  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

lrtest(effect\_cat,effect\_cat\_no)

## Likelihood ratio test  
##   
## Model 1: cvd ~ as.factor(age) + asp + as.factor(age) \* asp  
## Model 2: cvd ~ as.factor(age) + asp  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 6 -4442.0   
## 2 4 -4444.4 -2 4.7833 0.09148 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Since the p-value of the likely-hood ratio is greater than .05 we do not have enough evidence to conclude the need to adjust for effect modification for age and asprine use.

1. **Describe the implications of these results for use of aspirin for primary prevention of CVD in women.**

summary(confound)

##   
## Call:  
## glm(formula = cvd ~ as.factor(age) + asp, family = binomial(link = "logit"),   
## data = whsa)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4028 -0.2467 -0.1686 -0.1609 2.9507   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.24692 0.06387 -66.495 <2e-16 \*\*\*  
## as.factor(age)2 0.86331 0.07699 11.214 <2e-16 \*\*\*  
## as.factor(age)3 1.77586 0.08162 21.759 <2e-16 \*\*\*  
## asp -0.09337 0.06461 -1.445 0.148   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9338.9 on 39875 degrees of freedom  
## Residual deviance: 8888.8 on 39872 degrees of freedom  
## AIC: 8896.8  
##   
## Number of Fisher Scoring iterations: 7

Looking at the model adjust for age group asp had a p-value of 0.148 thus there is not enough evidence to conclude that asprin can significantly reduce CVD in woman.