BIOSTAT HW4

Ivy Zhang

10/27/2021

1. Based on a logistic regression model using grouped linear adjustment for age, the odds of getting esophageal cancer for patients who exposed to cider (> 10g/day) is 2.187 times (CI: 1.56-3.09, Wald Test p <0.001) the odds of patients who are in the same age group but not exposed to cider (<= 10g/day). This analysis provides evidence that cider exposure leads to higher odds of getting esophageal cancer.

2. Based on a logistic regression model using indicator variables adjusted for age, the odds of getting esophageal cancer for patients who exposed to cider (>10g/day) is 2.027 times (CI:1.45-2.863,Wald P-value <0.001) the odds of patients who are in the same age group but not exposed to cider (<=10g/day). This analysis provides evidence that cider exposure leads to higher odds of getting esophageal cancer.

3.(a) The estimated coefficients of these two analysis are similar(2.027 vs. 2.187). They are both statistically significant at the level of 0.05 using the Wald Test.

3.(b) I prefer to report the analysis using indicator variables to adjust for the age, it can provide more flexible adjustment of age groups by scarifising some degrees of freedom.

4.Based on a logistic regression model using continuous linear adjustment for age, the odds of getting esophaegeal cancer for patients who exposed to cider (>10g/day) is 2.168 times (CI: 1.549-3.057, Wald P-value < 0.001) the odds of patients who are at the same age but not exposed to cider (<=10g/day). This analysis provides evidence that cider exposure leads to higher odds of getting esophageal cancer.

5. Based on a logistic regression model using continuous quardratic adjustment for age, the odds of getting esophaegeal cancer for patients who exposed to cider(>10g/day) is 1.97 times (CI:1.40-2.78, Wald P-value < 0.001) the odds of patients who are at the same age but not exposed to cider (<=10g/day). The analysis provides evidence that cider exposure leads to higher odds of getting esophageal cancer.

6(a). We divided participants's age into 6 categories, where 35, 45, 55, 65, 75 are the cutpoints between the categories. Within each age intereval, we will using logistic regression to help us determining how the linear relationship between age in that interval changed from the last age interval (except for the first one). After adjusted the relationship between age and disease, we can estimate the effect of our exposure.

6(b). Based on a logistic regression model using linear spline adjusted for age, the odds of getting esophaegeal cancer for patients who exposed to cider (>10g/day) is 1.98 times (CI:1.41-2.798, Wald P-value:<0.001) the odds of patients who are at the same age but not exposed to cider (<=10g/day). The analysis provides evidence that cider exposures leads to higher odds of getting esophageal cancer.

7(a).I thought they are very similar. They are about 2.

7(b).I would prefer to report the results from the logistic regression model using linear spline adjusted for age. It can give more flexibility for adjusting for the age.

8(a).I think Q1 and Q2 are not nested.

8(b).I think Q1 and Q4 are not nested.

8(c).I think Q4 and Q5 are nested. The reduced model is the Q4 model. Q5 has an additionai I(age^2) compared to the Q4 model.

8(d).I think Q4 and Q6 are nested. The reduced model is the Q4 model. Q6 has additional spline variables compared to the Q4 model.

8(e).I think Q5 and Q6 are not nested.

Code Appendix

```
#----------------Set Up--------------------
knitr::opts_chunk$set(echo =FALSE,results= FALSE, warning = FALSE, message = FALSE,
                      fig.height =5)
library(dplyr)
library(dagR)
library(uwIntroStats)
library(reshape)
library(knitr)
options(digits = 3)
load("/Users/ivyyuezhang/Desktop/R hw/esophcts.Rdata")
esophcts$binci = 1
esophcts$binci[which(esophcts$cider<=10)] = 0
#-----------------Question 1-----------------------
log_group = glm(case~agegp + binci, data = esophcts, family = "binomial")
summary(log_group)
exp(coef(log_group))
orci = exp(confint(log_group))
orci
#----------------Question 2 ----------------------------
log_dummy = glm(case~as.factor(agegp)+binci, family = "binomial",data = esophcts)
summary(log_dummy)
exp(coef(log_dummy))
exp(confint(log_dummy))

#--------------------------Question 4 --------------------
log_con = glm(case~age+binci, data = esophcts, family = "binomial")
summary(log_con)
exp(coef(log_con))
exp(confint(log_con))
#------------------Question 5-------------------
log_con_qua = glm(case~age + binci + I(age^2),data = esophcts, family = "binomial")
summary(log_con_qua)
exp(coef(log_con_qua))
exp(confint(log_con_qua))
#----------------Question 6----------------
knots  = c(35,45,55,65,75)
for(k in 1:length(knots)){
  esophcts[,paste("spline",k,sep = "")] = esophcts$age-knots[k]
  esophcts[,paste("spline",
                  k,sep = "")][esophcts[,paste("spline",
                                               k,sep = "")]<0] = 0
}
log_spl = glm(case~age+spline1+spline2+spline3+spline4+spline5+binci,
              data = esophcts, family = "binomial")
summary(log_spl)
exp(coef(log_spl))
```

```r
exp(confint(log_spl))
```