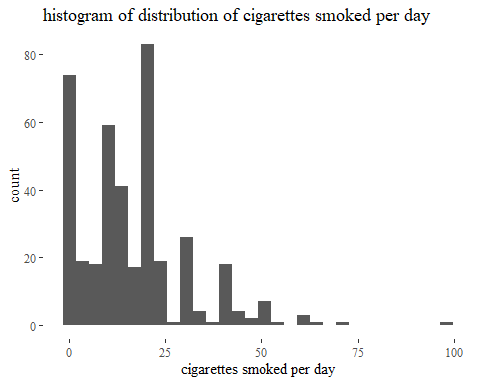
Brents hw8

Colleen Brents

April 3, 2019



Summary statistics of cigarettes/day by case status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| case | mean | median | max | min |
| 0 | 16 | 15 | 98 | 0 |
| 1 | 19 | 20 | 70 | 0 |

The amount of cigarettes per day among those who did not have esophageal cancer was less than that of those with esophageal cancer.

# 2.

In a case control study of esophageal cancer among 400 men from Singapore, we have evidence to reject the null hypothesis that hot beverage consumption is not associated with esophageal cancer (p=1e-06). The estimated odds of esophageal cancer for those who consumed hot beverages is 3.81 (95% confidence interval 2.21, 6.57).

# 3.

In a case control study of esophageal cancer among 400 men from Singapore, we have evidence to reject the null hypothesis that there is not a linear trend in hot beverage consumption and esophageal cancer hot beverage consumption (p<2e-16). The estimated of odds of esophageal cancer is2.31 (95% confidence interval 1.7, 3.13) times higher for a one unit difference in hot beverage consumption.

# 4.

In a case control study of esophageal cancer among 400 men from Singapore, we have evidence to reject the null hypothesis that there is no evidence of heterogeneity in the trend for higher amounts of beverage consumption and esophageal cancer (p< 0.05). The estimated odds are 2.25 (95% CI 1.05, 4.8) times higher for a man who consumes one hot beverage compared with none, adjusted for the same cigarette and sampu usage. The estimated odds are 4.41 (95% CI 2.1, 9.29) times higher for a man who consumes two hot beverages compared to none, among those with same sampu and cigarette intake. The estimated odds are 21.69 (95% CI 4.32, 108.86) times higher for a man who consumes three hot beverages compared to none, adjusted for those with the same sampu and cigarette consumption.

|  |  |  |  |
| --- | --- | --- | --- |
| Number of Hot Beverages | Binary est. OR | Continuous est. OR | Categorical est. OR |
| 0 | 1.0 | 1.0 | 1.0 |
| 1 | 3.8 | 2.3 | 2.2 |
| 2 | 3.8 | 5.3 | 4.4 |
| 3 | 3.8 | 12.2 | 21.7 |

## Appendix

knitr::opts\_chunk$set(  
 echo = FALSE,  
 message = FALSE,  
 warning = FALSE  
)  
options(scipen=1, digits=2)  
library(readr)  
library(ggplot2)  
library(splines)  
library(dplyr)  
library(tidyr)  
library(ggthemes)  
library(knitr)  
library(broom) #tidy   
library(stringr)  
library(lmtest)  
library(sandwich)  
library(car)  
library(survival)  
url <- paste0("https://raw.githubusercontent.com/cbrents/stat796logreg/master/hw8/singapore.csv")  
tea<-read\_csv(url)  
get\_ests\_table <- function(model\_obj){  
 model\_ests <- data.frame(names(coef(model\_obj)),  
 round(coef(model\_obj), digits=2),  
 round(sqrt(diag(sandwich(model\_obj))), digits=2),  
 round(exp(coef(model\_obj)), 2),  
 exp(confint(model\_obj)))  
model\_ests$ci <- paste0("(",  
 round(model\_ests[, 5], 2),  
 ",",  
 round(model\_ests[, 6], 2),  
 ")")  
model\_ests <- model\_ests[, -5:-6]  
colnames(model\_ests) <- c("Model Term", "Estimate", "SE",  
 "Exp(Est.)","95% CI")  
return(model\_ests)  
}  
ggplot(tea) +  
 geom\_histogram(aes(x=cigs\_day)) +  
 theme\_tufte() +  
 xlab("cigarettes smoked per day")+  
 ggtitle("histogram of distribution of cigarettes smoked per day")  
sum\_table<- tea %>%  
 group\_by(case) %>%   
 summarize('mean' = mean(cigs\_day),  
 'median' = median(cigs\_day),  
 'max' = max(cigs\_day),  
 'min' = min(cigs\_day))  
kable(sum\_table, caption="Summary statistics of cigarettes/day by case status")  
tea <- tea %>%   
 mutate (hotbev\_cat=as.factor(hotbev),  
 hot\_bevbin=hotbev,   
 hot\_bevbin=as.factor(hot\_bevbin))  
tea<- tea %>% mutate(hot\_bevbin=hotbev)  
tea$hot\_bevbin[tea$hot\_bevbin==1] <- "1"   
tea$hot\_bevbin[tea$hot\_bevbin==2] <- "1"  
tea$hot\_bevbin[tea$hot\_bevbin==3] <- "1"  
#tea$hot\_bevbin[tea$hot\_bevbin==2]<- "0"  
range(tea$cigs\_day) #0 to 98   
str(tea$cigs\_day\_gp) #20 levels  
str(tea$sampu) #binary   
 #treats hotbev binary   
#recode anew column treating 123 same   
tea\_cdl\_bin<- clogit(case~ hot\_bevbin +sampu+cigs\_day+strata(set), data=tea)  
summary(tea\_cdl\_bin) #0 is intercept (1?)   
tea\_cdl<- clogit(case~sampu+cigs\_day+strata(set), data=tea)  
anova(tea\_cdl, tea\_cdl\_bin, test="LRT")  
bin\_or<- exp(coef(tea\_cdl\_bin)[1])  
bin\_ci<-exp(confint(tea\_cdl\_bin, level=0.95))[1,1:2]  
# will be same OR for cups 1-3   
#hotbev continuous  
tea\_cdl\_con<- clogit(case~hotbev+sampu+cigs\_day+strata(set), data=tea)  
anova(tea\_cdl\_con, tea\_cdl\_bin, test="LRT")  
cont\_or<- exp(coef(tea\_cdl\_con)[1])  
cont\_ci<-exp(confint(tea\_cdl\_con, level=0.95))[1,1:2]  
#or;   
cont\_1<- exp(1\*coef(tea\_cdl\_con)[1]) #(1))#for one cup;   
cont\_2<- exp(2\*coef(tea\_cdl\_con)[1]) #2  
cont\_3<- exp(3\*coef(tea\_cdl\_con)[1]) #3   
#hotbev categorical   
tea\_cdl\_cat<- clogit(case~factor(hotbev)+sampu+cigs\_day+strata(set), data=tea)  
summary(tea\_cdl\_cat)  
anova(tea\_cdl\_con, tea\_cdl\_cat, test="LRT")  
cat\_or<- exp(coef(tea\_cdl\_cat))[1:3]  
cat\_1or<- exp(coef(tea\_cdl\_cat))[1]  
cat\_2or<- exp(coef(tea\_cdl\_cat))[2]  
cat\_3or<- exp(coef(tea\_cdl\_cat))[3]  
cat\_1ci<- exp(confint(tea\_cdl\_cat, level=0.95))[1,]  
cat\_2ci<- exp(confint(tea\_cdl\_cat, level=0.95))[2,]  
cat\_3ci<- exp(confint(tea\_cdl\_cat, level=0.95))[3,]  
cat\_ci<-exp(confint(tea\_cdl\_cat, level=0.95))[1:3,]  
num<- c(0:3)  
bina<-c(1, bin\_or, bin\_or, bin\_or)   
cont<- c(1, cont\_1, cont\_2, cont\_3)  
cata<- c(1, cat\_1or, cat\_2or, cat\_3or)  
table<- data.frame(num, bina, cont, cata)  
kable(table, col.names=c("Number of Hot Beverages", "Binary est. OR", "Continuous est. OR", "Categorical est. OR"), row.names = FALSE)