models\_for\_table\_2.R

81983

2019-12-14

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## <U+2713> ggplot2 3.2.1 <U+2713> purrr 0.3.3  
## <U+2713> tibble 2.1.3 <U+2713> dplyr 0.8.3  
## <U+2713> tidyr 1.0.0 <U+2713> stringr 1.4.0  
## <U+2713> readr 1.3.1 <U+2713> forcats 0.4.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(survey)

## Loading required package: grid

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

## Loading required package: survival

##   
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':  
##   
## dotchart

#### process data   
raw\_df = read\_rds("data/primary\_df\_1.rds")  
raw\_df = raw\_df %>% mutate(  
 smoke\_mod =   
 case\_when(  
 smoke100 == 1 ~"yes",  
 smoke100 == 2 ~"no",  
 TRUE ~ NA\_character\_  
 ),  
 cancer\_num =   
 ifelse(cancer == "Yes",1,0),  
   
 exercise\_cat =   
 case\_when(  
 exerany2 == 1 ~"yes",  
 exerany2 == 2 ~"no",  
 TRUE ~ NA\_character\_  
 ),  
   
 health\_cover =   
 case\_when(  
 hlthpln1 == 1 ~"yes",  
 hlthpln1 == 2 ~"no",  
 TRUE ~ NA\_character\_  
 ),  
 edu\_cat\_mod =   
 case\_when(  
 educa == 1 ~" Less than HS",  
 educa == 2 ~" Less than HS",  
 educa == 3 ~" Less than HS",  
 educa == 4 ~" HS graduate",  
 educa == 5 ~" Some College",  
 educa == 6 ~"College Graduate",  
 TRUE ~ NA\_character\_  
 ),  
 alcweek5\_mod = alcday5\_mod/7 # change data in to the "week" unit  
)  
raw\_df$edu\_cat\_mod %>% table()

## .  
## Less than HS HS graduate Some College College Graduate   
## 32327 118526 119472 163633

raw\_df$cancer = as.factor(raw\_df$cancer)  
raw\_df$educa = as.factor(raw\_df$educa)  
raw\_df$x.incomg = as.factor(raw\_df$x.incomg)  
# remove education unknown  
raw\_df$educa[raw\_df$educa == 9] = NA  
raw\_df$educa = factor(raw\_df$educa)  
  
  
#### Model part   
  
svy\_obj\_2 <- svydesign(id=~x.psu, weights= ~x.llcpwt, strata = ~x.state , data=raw\_df, nest=TRUE)  
glm\_model\_4 = svyglm(cancer ~ age\_l + sex1 + x.imprace + sleptim1\_cat + alcweek5\_mod + smoke\_mod + health\_cover + exercise\_cat + x.incomg + edu\_cat\_mod , design = svy\_obj\_2, family = "binomial")

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

confint\_glm <- function(object, parm, level = 0.95, ...) {  
 coef = coef(summary(object)) %>% as.data.frame()  
 coef\_CI = object %>% confint() %>% as.data.frame()  
 table = cbind(coef, coef\_CI) %>%   
 mutate(Exp.Est = round(exp(Estimate),4),  
 CIL = round(`2.5 %`,4),  
 CIU = round(`97.5 %`,4),  
 Std.Error = round(`Std. Error`,4),  
 Estimate = round(Estimate,4),  
 p.value = round(`Pr(>|t|)`,4)) %>%  
 dplyr::select(Estimate, Exp.Est, Std.Error, CIL, CIU, p.value)  
 rownames(table) <- rownames(coef)  
 return(table)  
}  
glm\_result\_1 = confint\_glm(glm\_model\_4)   
  
  
#### add variable names  
glm\_result\_1 =   
 glm\_result\_1 %>%  
 rownames\_to\_column() %>%  
 mutate(  
 rownames\_new =  
 case\_when(  
 rowname == "(Intercept)" ~ "Intercept",  
 rowname == "age\_l25-34" ~ "Age (25-34)",  
 rowname == "age\_l35-44" ~ "Age (35-44)",  
 rowname == "age\_l45-54" ~ "Age (45-54)",  
 rowname == "age\_l55-64" ~ "Age (55-64)",  
 rowname == "age\_l65+" ~ "Age (65+)",  
 rowname == "sex1Male" ~ "Sex (Male)",  
 rowname == "x.impraceNon-Hispanic, Black" ~ "Non-Hispanic, Black",  
 rowname == "x.impraceNon-Hispanic, Other" ~ "Non-Hispanic, Other",  
 rowname == "x.impraceNon-Hispanic, White" ~ "Non-Hispanic, White",  
 rowname == "sleptim1\_cat2 Adquate sleep" ~ "Sleep (Adequate)",  
 rowname == "sleptim1\_cat3 Excessive sleep" ~ "Sleep (Excessive)",  
 rowname == "alcweek5\_mod" ~ "Weeks of drinks",  
 rowname == "smoke\_modyes" ~ "Smoke Status",  
 rowname == "health\_coveryes" ~ "Health Coverage",  
 rowname == "exercise\_catyes" ~ "Exercise Status",  
 rowname == "x.incomg2" ~ "$15,000 ~ $25,000",  
 rowname == "x.incomg3" ~ "$25,000 ~ $35,000",  
 rowname == "x.incomg4" ~ "$35,000 ~ $50,000",  
 rowname == "x.incomg5" ~ "More than $50,000 ",  
 rowname == "x.incomg9" ~ "unknown",  
 # rowname == "educa2" ~ "Elementary",  
 # rowname == "educa3" ~ "Some high school",  
 # rowname == "educa4" ~ "High school graduate",  
 # rowname == "educa5" ~ "Some college",  
 # rowname == "educa6" ~ "College graduate",  
 rowname == "edu\_cat\_mod HS graduate" ~ "High School Graduate",  
 rowname == "edu\_cat\_mod Some College" ~ "Some College",  
 rowname == "edu\_cat\_modCollege Graduate" ~ "College Graduate",  
 TRUE ~ NA\_character\_  
 )  
 ) %>%  
 remove\_rownames() %>%  
 dplyr::select(-rowname) %>%  
 column\_to\_rownames(var = "rownames\_new")  
  
#### print result   
glm\_result\_1 %>% knitr::kable()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | Exp.Est | Std.Error | CIL | CIU | p.value |
| Intercept | -5.3769 | 0.0046 | 0.1324 | -5.6365 | -5.1173 | 0.0000 |
| Age (25-34) | 0.7235 | 2.0615 | 0.1162 | 0.4956 | 0.9513 | 0.0000 |
| Age (35-44) | 1.4571 | 4.2936 | 0.1115 | 1.2386 | 1.6756 | 0.0000 |
| Age (45-54) | 2.2183 | 9.1918 | 0.1087 | 2.0052 | 2.4315 | 0.0000 |
| Age (55-64) | 2.7953 | 16.3672 | 0.1066 | 2.5863 | 3.0043 | 0.0000 |
| Age (65+) | 3.5821 | 35.9492 | 0.1056 | 3.3752 | 3.7890 | 0.0000 |
| Sex (Male) | -0.1688 | 0.8447 | 0.0197 | -0.2074 | -0.1301 | 0.0000 |
| Non-Hispanic, Black | -0.0828 | 0.9206 | 0.0670 | -0.2142 | 0.0486 | 0.2170 |
| Non-Hispanic, Other | 0.0172 | 1.0174 | 0.0761 | -0.1319 | 0.1664 | 0.8210 |
| Non-Hispanic, White | 0.7659 | 2.1508 | 0.0514 | 0.6650 | 0.8667 | 0.0000 |
| Sleep (Adequate) | -0.0636 | 0.9383 | 0.0212 | -0.1052 | -0.0220 | 0.0027 |
| Sleep (Excessive) | 0.1629 | 1.1770 | 0.0845 | -0.0027 | 0.3286 | 0.0539 |
| Weeks of drinks | 0.0287 | 1.0291 | 0.0078 | 0.0135 | 0.0439 | 0.0002 |
| Smoke Status | 0.1656 | 1.1801 | 0.0199 | 0.1265 | 0.2046 | 0.0000 |
| Health Coverage | 0.4303 | 1.5377 | 0.0541 | 0.3242 | 0.5363 | 0.0000 |
| Exercise Status | -0.1025 | 0.9026 | 0.0229 | -0.1473 | -0.0577 | 0.0000 |
| $15,000 ~ $25,000 | -0.0461 | 0.9550 | 0.0473 | -0.1388 | 0.0467 | 0.3304 |
| $25,000 ~ $35,000 | -0.0580 | 0.9437 | 0.0515 | -0.1590 | 0.0430 | 0.2607 |
| $35,000 ~ $50,000 | -0.0334 | 0.9671 | 0.0487 | -0.1289 | 0.0620 | 0.4922 |
| More than $50,000 | -0.0911 | 0.9129 | 0.0441 | -0.1775 | -0.0048 | 0.0386 |
| unknown | -0.1002 | 0.9047 | 0.0457 | -0.1898 | -0.0106 | 0.0284 |
| High School Graduate | 0.0072 | 1.0072 | 0.0449 | -0.0809 | 0.0952 | 0.8734 |
| Some College | 0.1446 | 1.1555 | 0.0456 | 0.0552 | 0.2340 | 0.0015 |
| College Graduate | 0.2335 | 1.2631 | 0.0459 | 0.1436 | 0.3235 | 0.0000 |