Smoking

Hank

## Summary of conclusions

On the one hand, smoking of cigars likelihood comparison among European Americans, Hispanic Americans and African Americans shows no adequate evidence of higher likelihood in European Americans given they are from the same areas (either rural or urban). On the other hand, gender plays a marginal part in determining the likelihood of having used a eletronic cigarette once individuals have similar age, ethinicity, grade and living area.

## Introduction and model

#### Hypothesis 1

The first hypothesis I tested is whether smoking of cigars, cigarillo or little cigars is more common amongst European Americans than Hispanic-Americans and African-Americans or not if they all live in the rural or urban area. And this can be tested by looking at the relation between *ever\_cigars\_cigarillos\_or* and the interaction term of *race* and *RuralUrban*. Note that, in the dataset, in addition to 3 races mentioned before, there are pacific, native, and asian which only accounts for a very small proportion of the data which cannot add any accuracy to my analysis, so I decided to exclude thoes observations which finally leaves 17298 observations in the dataset for use.

The model is:

and the hypothesis I want to test is:

where indicates whether this observation ever smoked cigars, cigarillos or little cigars. The interaction term *UrbanRural:Race* in the regression allows me to distinguish the impact of ethinicities living in the same area.

Table1: Cigar smoking for different races

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -1.852 | 0.061 | -30.556 | 0.000 |
| Racewhite:RuralUrbanUrban | -0.259 | 0.078 | -3.338 | 0.001 |
| Raceblack:RuralUrbanUrban | 0.044 | 0.097 | 0.455 | 0.649 |
| Racehispanic:RuralUrbanUrban | -0.198 | 0.082 | -2.408 | 0.016 |
| Racewhite:RuralUrbanRural | 0.217 | 0.072 | 3.025 | 0.002 |
| Raceblack:RuralUrbanRural | 0.686 | 0.096 | 7.154 | 0.000 |

The regression estimates and significance confirms that smoking of cigars is not more common amongst European Americans than Hispanic-Americans and African-Americans. To be exact, European Americans have the lowest likelihood of smoking cigars in urban area, while in Rural area, European Americans still have lower likelihood than Afriance-Americans to smoke.

#### research2

The second hypothesis I tested is whether gender has any impact on likelihood of having used a electronic cigarette given if individuals have similar demographic characteristics like *Age*, *Grade*, *Race*, *RuralUrban*.

The model is:

And the test I want to test is:

where indicates whether the individual ever used a electronic cigarette. The estimated coefficient of *Sex* and its p-value allows me to determine the impact of *Sex* on likelihood of ever used a electronic cigarette.

Table2: Likelihood of ever used eletronic cigarette

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -4.968 | 0.187 | -26.626 | 0.000 |
| SexF | -0.058 | 0.034 | -1.713 | 0.087 |
| Age | 0.218 | 0.028 | 7.785 | 0.000 |
| Grade | 0.131 | 0.029 | 4.523 | 0.000 |
| Raceblack | -0.502 | 0.054 | -9.365 | 0.000 |
| Racehispanic | -0.091 | 0.038 | -2.397 | 0.017 |
| RuralUrbanRural | 0.148 | 0.034 | 4.328 | 0.000 |

From the table above, we can see that, the p-value of *Sex* is which is larger than the significance level which means it does not affect the likelihood of smoking significantly given all else variables are the same. In conclusion, two individuals of the different sexes are equally likely to ever used a electronic cigarette given they have similar *Age*, *Grade*, *Race* and *RuralUrban*.

#### Appendix: code

dataDir = "C:/desktop"  
smokeFile = file.path(dataDir, "smokeDownload.RData")  
  
if (!file.exists(smokeFile)) {  
download.file("http://pbrown.ca/teaching/appliedstats/data/smoke.RData", smokeFile)  
}else(  
load(smokeFile)  
)  
  
smokeSub = smoke[which(smoke$Age >= 10), ]  
smokeSub = data.frame(smokeSub)  
  
smokeSub = smokeSub[,c("Sex","Grade","Age","Race","RuralUrban","ever\_ecigarette","ever\_cigars\_cigarillos\_or")]  
  
##research 1  
smokeSub1 = smokeSub[smokeSub$Race %in% c('white','hispanic',"black"),]  
  
res1 = glm(ever\_cigars\_cigarillos\_or ~ Race:RuralUrban,data = smokeSub1,family = binomial(link = "logit") )  
  
knitr::kable(summary(res1)$coef,digits = 3,caption = 'Table1: Cigar smoking for different races',table.envir = "table\*")  
  
## research 2  
res2 = glm(ever\_ecigarette ~ Sex+ Age + Grade + Race + RuralUrban ,data = smokeSub1 ,family = binomial(link = "logit"))  
  
knitr::kable(summary(res2)$coef,digits = 3, caption = 'Table2: Likelihood of ever used eletronic cigarette')