Question 1

One hundred twenty-five male fruit flies were divided into five separate groups. Each group contained different numbers of female, either virgin or pregnant. By controlling the thorax length of each male (which was known to affect lifetime), the lifetime (in days) of each male fruit fly was measured. This experiment aimed to research how the longevity of fruit flies (in days) depending on sexual activity and thorax length.

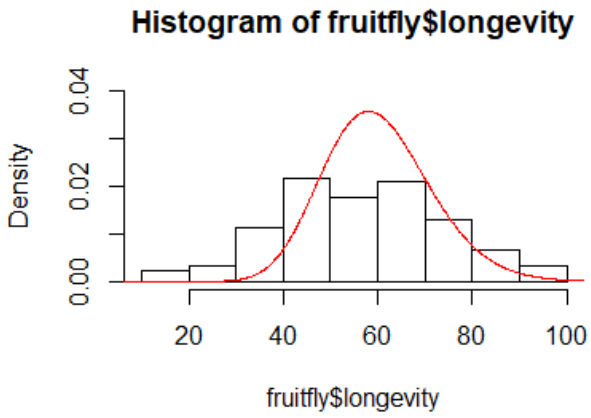
We fit a gamma regression model to control for the effects of thorax size. If we fixed the thorax of fruit flies, the fly kept with eight pregnant fruit flies had the most extended lifetime and the fly kept with eight virgin fruit flies had the shortest lifetime. The table indicated that the virgin fruit fly (e.g., activity low and activity high) had a negative influence on male fruit fly's lifetime. Moreover, the pregnant fruit fly had a positive impact on male fruit fly's lifetime. Furthermore, the long thorax length caused a longer lifetime.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	4.098	0.038	108.333	0.000
norm_thorax	0.204	0.017	11.804	0.000
fruitfly$activityone	0.055	0.053	1.036	0.302
fruitfly$activitylow	-0.116	0.053	-2.184	0.031
fruitfly$activitymany	0.082	0.054	1.524	0.130
fruitfly$activityhigh	-0.415	0.054	-7.687	0.000

Summary table: estimated parameters from the Gamma generalized linear model of fruit fly

(NOTE: The group: isolated = fly kept solitary, one = fly kept with one pregnant fruitfly, many = fly kept with eight pregnant fruitflies, low= fly kept with one virgin fruitfly, high = fly kept with eight virgin fruitflies.)

Finally, we used histogram of data to check if the data fits the model well. The data and fitted line followed the same pattern (looks gamma) which was good.



Question 2

Summary

We analyzed the data from the 2014 American National Youth Tobacco Survey to find out the parameters that influenced the odds of regular use of chewing tobacco and the odds of using a hookah (or water pipe). We wanted to research how the race, age, sex, and living place affected human behavior. The data indicated that the white Americans who lived in the rural area more likely to chew tobacco than Hispanic and African (black people). Hispanic and African were approximate 0.5 and 0.2 of the odd rates separately. Furthermore, race pacific had the most significant proportion of chewing tobacco compared to others. Overall, aged males who lived in rural areas had the most significant odds on regular use of chewing tobacco. Also, if kept race, age and demographic background similar, the likelihood of having used a hookah or waterpipe once before was not influenced by the smoker's sex. Moreover, aged rural people more likely to use hookah or waterpipe. In a surprise, the African-American (black people) had less tendency to use hookah or waterpipe followed by Asian.

Introduction

We used R to analyze the 2014 American National Youth Tobacco Survey data in detail. The data was available from http://pbrown.ca/. We focus on the paraments that could affect the use of cigars, hookahs, and chewing tobacco. For the first model, we mainly research the correlation between the odds of regular use of chewing tobacco and the races. (Note: Regular use of tobacco products, defined as use on ≥1 day during the past 30 days.) If applicable, we also could explore other factors' (age & sex& living place) influence on chewing tobacco. For the second model, we wanted to know if the gender had effects on the likelihood of having used a hookah or waterpipe on at least one occasion. For both models, we controlled for age, living place (rural or urban), and other demographic characteristics of smokers similar.

Methods

For analysis, we used a logistic generalized linear model that fits the data since y is 1,0. In this case, consider the following model:

$$\ln(Odds) = \beta_0 + \beta_1 x_{ageC} + \beta_2 I_{SexF} + \beta_3 I_{RaceBlack(\text{African})} + \beta_4 I_{RaceHispanic} + \beta_5 I_{RaceAsian} + \beta_6 I_{RaceNative} + \beta_7 I_{RacePacific} + \beta_8 I_{RuralUrban}$$

$$Odds = \frac{\pi}{1-\pi} \quad larger\ odds, larger\ probability$$

In the first model, Odds used to calculate the probability of Regular use of Chewing tobacco, snuff or dip. We tested the significance of race on odds; therefore, the hypothesis test was: $H_0: \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$.

Moreover, in the second model, Odds used to calculate the probability of the usage of hookah .We tested the significance of sex on odds; therefore, the hypothesis test was: $H_0: \beta_2 = 0$.

Confidence Intervals for the Odds Ratio: $e^{\hat{\beta} \pm SE\,\hat{\beta}}$

(Noticed that, we may not use Race Asian, Native, and Pacific to explore our main topic, but we kept it

for further information.)

Results

First Model: Race & Chewing tobacco

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	-3.032	0.083	-36.483	0.000
ageC	0.337	0.021	16.204	0.000
SexF	-1.788	0.109	-16.481	0.000
Raceblack	-1.556	0.172	-9.064	0.000
Racehispanic	-0.713	0.104	-6.884	0.000
Raceasian	-1.546	0.342	-4.519	0.000
Racenative	0.107	0.278	0.385	0.700
Racepacific	1.012	0.361	2.807	0.005
RuralUrbanRural	0.951	0.087	10.876	0.000

Summary table 1: The log-odds of regular use of chewing tobacco

From the Summary table 1, the p-value of Race black, Race white and Race Hispanic are smaller than 0.05. In other words, all those parameters are significant to the regular use of chewing tobacco.

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	0.048	1.087	0.000	1.000
ageC	1.400	1.021	10900445.045	1.000
SexF	0.167	1.115	0.000	1.000
Raceblack	0.211	1.187	0.000	1.000
Racehispanic	0.490	1.109	0.001	1.000
Raceasian	0.213	1.408	0.011	1.000
Racenative	1.113	1.320	1.470	2.014
Racepacific	2.751	1.434	16.563	1.005
RuralUrbanRural	2.588	1.091	52876.794	1.000

Summary table 2: The odds of regular use of chewing tobacco

Then, we took the exponential to the summary table and checked how the exponential coefficients change the Odds. After we control for age and sex, we notice that white Americans lived in Rural Area are more likely to chew tobacco. The Hispanic-American and African-American(black) are about 20% and 40% less likely to chew tobacco than white. Furthermore, we can see that women chew tobacco are 17% to men. Also, the rural area residents are about 2.58 times likely to chew tobacco compared to the Urban citizens.

	exp_lower_bound	exp_upper_bound
(Intercept)	0.04083615	0.05693957
ageC	1.34314320	1.45949344
SexF	0.13460476	0.20776153
Raceblack	0.14965447	0.29738045
Racehispanic	0.39840997	0.60294051
Raceasian	0.10747078	0.42232793
Racenative	0.63879725	1.93876768
Racepacific	1.33780215	5.65872158
RuralUrbanRural	2.17264417	3.08218339

Summary table 3: The Confidence Interval of the Odds

Second Model: Sex & Using a Hookah

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	-1.724	0.044	-39.226	0.000
ageC	0.419	0.012	36.266	0.000
SexF	0.042	0.043	0.980	0.327
Raceblack	-0.635	0.070	-9.005	0.000
Racehispanic	0.346	0.048	7.138	0.000
Raceasian	-0.631	0.118	-5.362	0.000
Racenative	0.160	0.190	0.838	0.402
Racepacific	0.964	0.270	3.566	0.000
RuralUrbanRural	-0.388	0.044	-8.769	0.000

Summary table 4: The log-odds of use a hookah

From the Summary table 4, the p-value of sex is larger than 0.05. In other words, sex is not significant to the use of hookah. Controlling for age, ethnicity and other demographic characteristics similar, the sex of smokers cannot influence their behavior on 'ever use a hookah.'

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.000	1.204	0.000000e+00	1.000
Age	1.520	1.012	5.623341e+15	1.000
SexF	1.043	1.044	2.666000e+00	1.387
Raceblack	0.530	1.073	0.000000e+00	1.000
Racehispanic	1.413	1.050	1.258299e+03	1.000
Raceasian	0.532	1.125	5.000000e-03	1.000
Racenative	1.173	1.210	2.312000e+00	1.495
Racepacific	2.621	1.310	3.538700e+01	1.000
RuralUrbanRural	0.678	1.045	0.000000e+00	1.000

Summary table 5: The odds of use of hookah

From the summary table 5, we noticed that race pacific is much preferred to use hookah compared to another race. A unit increase in age is associated with a 52% increase in odds. Surprisingly, the African-America(black) are less likely than white to try a hookah. This result is reversely comparing the result in the first model. Also, the rural residents are almost as same likely to try hookah as urban. Moreover, the table indicates that the odds of use of hookah of women are 4% higher than men. Since sex is not a statistically significant parameter, we only can conclude that the likelihood of having used a hookah on at least one occasion is the same for two individuals of different sex (, provided other background information same.)

	exp_lower_bound	exp_upper_bound
(Intercept)	0.1634291	0.1948306
ageC	1.4852154	1.5554041
SexF	0.9572099	1.1363689
Raceblack	0.4603920	0.6103322
Racehispanic	1.2824617	1.5565615
Raceasian	0.4205311	0.6733007
Racenative	0.8015719	1.7167054
Racepacific	1.5268970	4.4999523
RuralUrbanRural	0.6206370	0.7409467

Summary table 6: The Confidence Interval of the Odds

Appendix

Q1

```
data('fruitfly', package='faraway')
summary(fruitfly)
```

```
##      thorax         longevity         activity
##  Min.   :0.6400   Min.   :16.00   isolated:25
##  1st Qu.:0.7600   1st Qu.:46.00   one     :25
##  Median :0.8400   Median :58.00   low     :25
##  Mean   :0.8224   Mean   :57.62   many    :24
##  3rd Qu.:0.8800   3rd Qu.:70.00   high    :25
##  Max.   :0.9400   Max.   :97.00
```

```
#normalize thorax
mth = mean(fruitfly$thorax)
sd = sd(fruitfly$thorax)

norm_thorax = (fruitfly$thorax-mth)/sd

mod_fruitfly = glm(fruitfly$longevity ~ norm_thorax + fruitfly$activity, family = Gamma(link=lo
g))
summary(mod_fruitfly)
```

```
##
## Call:
## glm(formula = fruitfly$longevity ~ norm_thorax + fruitfly$activity,
##     family = Gamma(link = log))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.50718  -0.15216  -0.02833   0.12434   0.39938
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.09771    0.03783 108.333  < 2e-16 ***
## norm_thorax            0.20433    0.01731  11.804  < 2e-16 ***
## fruitfly$activityone   0.05527    0.05337   1.036   0.3024
## fruitfly$activitylow  -0.11646    0.05332  -2.184   0.0309 *
## fruitfly$activitymany  0.08250    0.05413   1.524   0.1302
## fruitfly$activityhigh -0.41466    0.05394  -7.687 4.93e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.0355297)
##
##     Null deviance: 13.2803  on 123  degrees of freedom
## Residual deviance:  4.3151  on 118  degrees of freedom
## AIC: 942.29
##
## Number of Fisher Scoring iterations: 4
```

```
knitr::kable(rbind(summary(mod_fruitfly)$coef), digits=3)
```

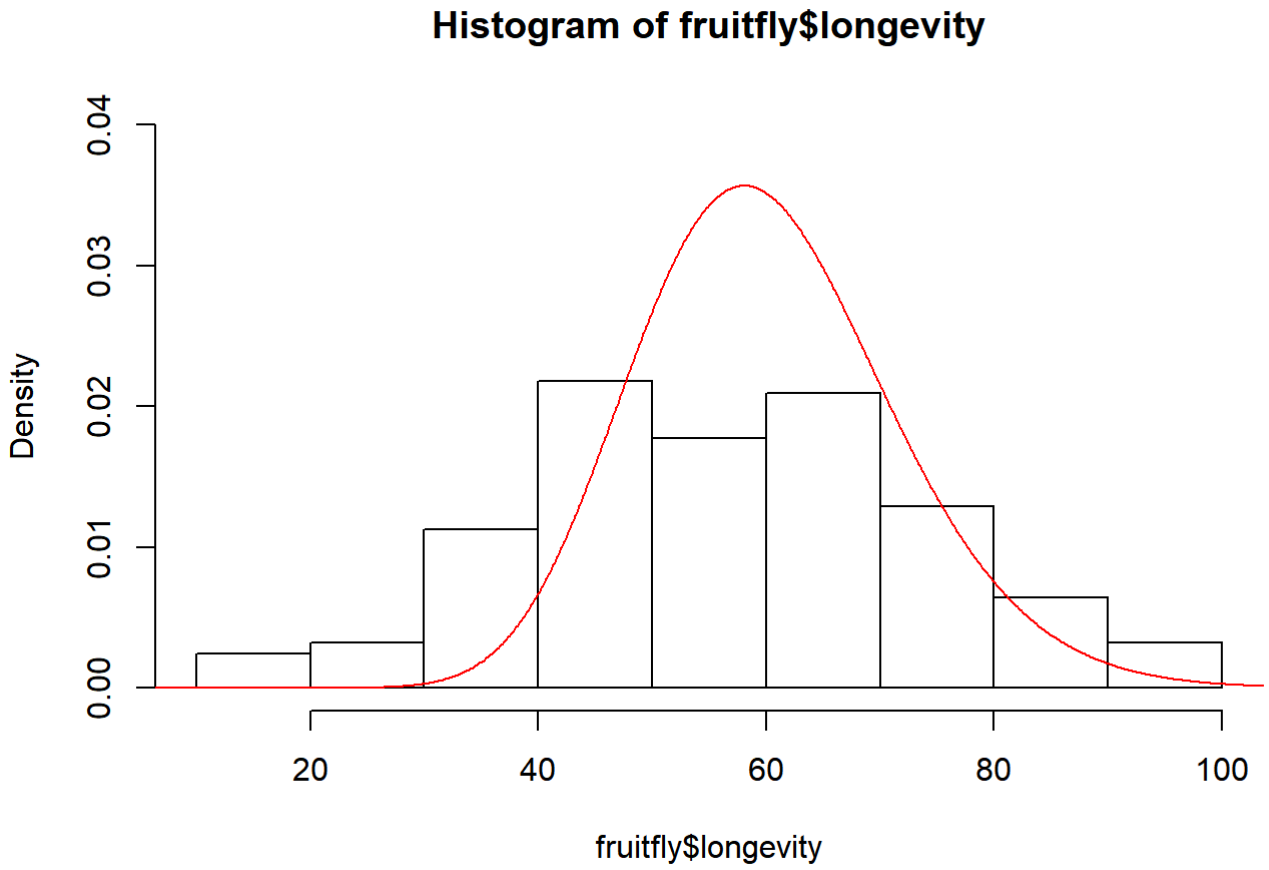
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.098	0.038	108.333	0.000
norm_thorax	0.204	0.017	11.804	0.000
fruitfly$activityone	0.055	0.053	1.036	0.302
fruitfly$activitylow	-0.116	0.053	-2.184	0.031
fruitfly$activitymany	0.082	0.054	1.524	0.130
fruitfly$activityhigh	-0.415	0.054	-7.687	0.000

```
# historgram
shape = 1/summary(mod_fruitfly)$dispersion

intercept_est=mod_fruitfly$coefficients[1]

scale = exp(intercept_est)/shape

hist(fruitfly$longevity,prob=TRUE,ylim=c(0,0.04))
xSeq=seq(0,120,len=1000)
lines(xSeq,dgamma(xSeq,shape = shape,scale = scale),col="red")
```


Histogram of fruitfly$longevity
Density
0.00
0.01
0.02
0.03
0.04
20
40
60
80
100
fruitfly$longevity

Question 2 - model 1

```
# First Model
smokeUrl = 'http://pbrown.ca/teaching/appliedstats/data/smoke.RData'
(smokeFile = tempfile(fileext='.RData'))
```

```
## [1] "C:\\Users\\WLJY8\\AppData\\Local\\Temp\\Rtmpkd4UBa\\filec241f105988.RData"
```

```
download.file(smokeUrl, smokeFile, mode='wb')
(load(smokeFile))
```

```
## [1] "smoke"        "smokeFormats"
```

```
smoke[1:20,c('Age','Sex','Grade','RuralUrban','Race', 'chewing_tobacco_snuff_or')]
```

```
##    Age Sex Grade RuralUrban     Race chewing_tobacco_snuff_or
## 1   13   M     2      Urban hispanic                    FALSE
## 2   12   F     2      Urban hispanic                    FALSE
## 3   14   M     2      Urban   native                    FALSE
## 4   13   M     2      Urban hispanic                    FALSE
## 5   14   M     2      Urban   native                    FALSE
## 6   13   F     3      Urban   native                     TRUE
## 7   14   M     3      Urban hispanic                    FALSE
## 8   14   F     3      Urban   native                    FALSE
## 9   14   F     3      Urban     <NA>                    FALSE
## 10  14   F     3      Urban   native                    FALSE
## 11  13   F     3      Urban   native                    FALSE
## 12  14   F     3      Urban hispanic                    FALSE
## 13  14   F     3      Urban hispanic                    FALSE
## 14  15   F     3      Urban     <NA>                     TRUE
## 15  14   M     3      Urban     <NA>                    FALSE
## 16  14   M     3      Urban   native                    FALSE
## 17  13   F     3      Urban   native                    FALSE
## 18  14   M     3      Urban hispanic                     TRUE
## 19  14   M     3      Urban   native                    FALSE
## 20  14   F     3      Urban   native                    FALSE
```

```
smokeFormats[smokeFormats$colName == 'chewing_tobacco_snuff_or', ]
```

```
##         ID
## 151 cslt_r
##                                                                                label
## 151 RECODE: Used chewing tobacco, snuff, or dip on 1 or more days in the past 30 days
##                                                          shortLabel
## 151 chewing tobacco snuff or dip on 1 or more days in the past 30 days
##                     colName
## 151 chewing_tobacco_snuff_or
```

```
smoke$everSmoke = factor(smoke$chewing_tobacco_snuff_or, levels=c('TRUE','FALSE'), labels=c('ye
s','no'))
table(smoke$Grade, smoke$Age, exclude=NULL)
```

```
##
##           9   10   11   12   13   14   15   16   17   18   19 <NA>
##   1      13    8 1311 1806  192    9    3    2    3    0    2    8
##   2       6    2   13 1267 2029  201   12    1    0    0    1    9
##   3       2    0    0    3 1379 1907  211   10    2    3    1    3
##   4       4    1    0    0    6 1085 1581  181   16    3    2    6
##   5       0    0    0    1    1   10 1114 1593  188   18    4    4
##   6       3    0    0    0    0    1    3 1089 1524  183   11    3
##   7      10    0    0    0    1    1    1   13 1109 1471  153    5
##   8       2    1    1    0    2    0    0    1    3    1    6    1
##   <NA>    0    0    2    5    7    4   14   10    8    3    0  118
```

```
table(smoke$Race, smoke$everSmoke, exclude=NULL)
```

```
##
##             yes   no <NA>
##   white     527 9300   66
##   black      40 3317   73
##   hispanic  145 5820  116
##   asian      10  954    9
##   native     15  320    3
##   pacific    11   71    3
##   <NA>       47 1054  106
```

```
smokeSub = smoke[smoke$Age >= 10 & !is.na(smoke$Race) &
                   !is.na(smoke$everSmoke) & !is.na(smoke$chewing_tobacco_snuff_or) &!is.na(smo
ke$Sex), ]
dim(smokeSub)
```

```
## [1] 20426   163
```

```
smokeAgg = reshape2::dcast(smokeSub,
                           Age + Sex + Race + RuralUrban ~ everSmoke,
                           length)
```

```
## Using everSmoke as value column: use value.var to override.
```

```
dim(smokeAgg)
```

```
## [1] 210   7
```

```
smokeAgg = na.omit(smokeAgg)
dim(smokeAgg)
```

```
## [1] 209   7
```

```
smokeAgg$y = cbind(smokeAgg$yes, smokeAgg$no)
smokeFit = glm(y ~ Age + Sex + Race + RuralUrban,
               family=binomial(link='logit'), data=smokeAgg)

knitr::kable(summary(smokeFit)$coef, digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.080	0.337	-23.997	0.000
Age	0.337	0.021	16.204	0.000
SexF	-1.788	0.109	-16.481	0.000
Raceblack	-1.556	0.172	-9.064	0.000
Racehispanic	-0.713	0.104	-6.884	0.000
Raceasian	-1.546	0.342	-4.519	0.000
Racenative	0.107	0.278	0.385	0.700
Racepacific	1.012	0.361	2.807	0.005
RuralUrbanRural	0.951	0.087	10.876	0.000

```
smokeAgg$ageC = smokeAgg$Age - 15
smokeFit2 = glm(y ~ ageC + Sex + Race + RuralUrban,
                family=binomial(link='logit'), data=smokeAgg)

knitr::kable(summary(smokeFit2)$coef, digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.032	0.083	-36.483	0.000
ageC	0.337	0.021	16.204	0.000
SexF	-1.788	0.109	-16.481	0.000
Raceblack	-1.556	0.172	-9.064	0.000
Racehispanic	-0.713	0.104	-6.884	0.000
Raceasian	-1.546	0.342	-4.519	0.000
Racenative	0.107	0.278	0.385	0.700
Racepacific	1.012	0.361	2.807	0.005
RuralUrbanRural	0.951	0.087	10.876	0.000

```
#The odds of regular use of chewing tobacco
knitr::kable(exp(summary(smokeFit2)$coef), digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.048	1.087	0.000	1.000

	Estimate	Std. Error	z value	Pr(>|z|)
ageC	1.400	1.021	10900445.045	1.000
SexF	0.167	1.115	0.000	1.000
Raceblack	0.211	1.187	0.000	1.000
Racehispanic	0.490	1.109	0.001	1.000
Raceasian	0.213	1.408	0.011	1.000
Racenative	1.113	1.320	1.470	2.014
Racepacific	2.751	1.434	16.563	1.005
RuralUrbanRural	2.588	1.091	52876.794	1.000

```
# The CI of parameter
sum =summary(smokeFit2)
est=sum$coefficients[,1]
std=sum$coefficients[,2]

exp_upper_bound = exp(est+2*std)
exp_upper_bound
```

```
##     (Intercept)            ageC            SexF       Raceblack
##      0.05693957      1.45949344      0.20776153      0.29738045
##    Racehispanic       Raceasian      Racenative     Racepacific
##      0.60294051      0.42232793      1.93876768      5.65872158
## RuralUrbanRural
##      3.08218339
```

```
exp_lower_bound= exp(est-2*std)
exp_lower_bound
```

```
##     (Intercept)            ageC            SexF       Raceblack
##      0.04083615      1.34314320      0.13460476      0.14965447
##    Racehispanic       Raceasian      Racenative     Racepacific
##      0.39840997      0.10747078      0.63879725      1.33780215
## RuralUrbanRural
##      2.17264417
```

```
cbind(exp_lower_bound,exp_upper_bound)
```

```
##                 exp_lower_bound exp_upper_bound
## (Intercept)          0.04083615      0.05693957
## ageC                 1.34314320      1.45949344
## SexF                 0.13460476      0.20776153
## Raceblack            0.14965447      0.29738045
## Racehispanic         0.39840997      0.60294051
## Raceasian            0.10747078      0.42232793
## Racenative           0.63879725      1.93876768
## Racepacific          1.33780215      5.65872158
## RuralUrbanRural      2.17264417      3.08218339
```

Question 2 - model 2

```
# Second Model
smokeUrl = 'http://pbrown.ca/teaching/appliedstats/data/smoke.RData'
(smokeFile = tempfile(fileext='.RData'))
```

```
## [1] "C:\\Users\\WLJY8\\AppData\\Local\\Temp\\RtmpKYkQqj\\file413861953e34.RData"
```

```
download.file(smokeUrl, smokeFile, mode='wb')
(load(smokeFile))
```

```
## [1] "smoke"        "smokeFormats"
```

```
dim(smoke)
```

```
## [1] 22007   162
```

```
smoke[1:20,c('Age','Sex','Grade','RuralUrban','Race', 'ever_tobacco_hookah_or_wa')]
```

```
##    Age Sex Grade RuralUrban     Race ever_tobacco_hookah_or_wa
## 1   13   M     2      Urban hispanic                     FALSE
## 2   12   F     2      Urban hispanic                     FALSE
## 3   14   M     2      Urban   native                     FALSE
## 4   13   M     2      Urban hispanic                      TRUE
## 5   14   M     2      Urban   native                     FALSE
## 6   13   F     3      Urban   native                     FALSE
## 7   14   M     3      Urban hispanic                     FALSE
## 8   14   F     3      Urban   native                     FALSE
## 9   14   F     3      Urban     <NA>                        NA
## 10  14   F     3      Urban   native                     FALSE
## 11  13   F     3      Urban   native                     FALSE
## 12  14   F     3      Urban hispanic                     FALSE
## 13  14   F     3      Urban hispanic                     FALSE
## 14  15   F     3      Urban     <NA>                        NA
## 15  14   M     3      Urban     <NA>                     FALSE
## 16  14   M     3      Urban   native                     FALSE
## 17  13   F     3      Urban   native                     FALSE
## 18  14   M     3      Urban hispanic                      TRUE
## 19  14   M     3      Urban   native                     FALSE
## 20  14   F     3      Urban   native                     FALSE
```

```
smokeFormats[smokeFormats$colName == 'ever_tobacco_hookah_or_wa', ]
```

```
##            ID                                                label
## 145 ehookah_r RECODE: Ever smoked tobacco out of a hookah or waterpipe
##                           shortLabel                   colName
## 145 ever tobacco hookah or waterpipe ever_tobacco_hookah_or_wa
```

```
smoke$everSmoke = factor(smoke$ever_tobacco_hookah_or_wa, levels=c('TRUE','FALSE'), labels=c('yes','no'))
table(smoke$Grade, smoke$Age, exclude=NULL)
```

```
##
##          9   10   11   12   13   14   15   16   17   18   19 <NA>
##   1     13    8 1311 1806  192    9    3    2    3    0    2    8
##   2      6    2   13 1267 2029  201   12    1    0    0    1    9
##   3      2    0    0    3 1379 1907  211   10    2    3    1    3
##   4      4    1    0    0    6 1085 1581  181   16    3    2    6
##   5      0    0    0    1    1   10 1114 1593  188   18    4    4
##   6      3    0    0    0    0    1    3 1089 1524  183   11    3
##   7     10    0    0    0    1    1    1   13 1109 1471  153    5
##   8      2    1    1    0    2    0    0    1    3    1    6    1
##   <NA>   0    0    2    5    7    4   14   10    8    3    0  118
```

```
table(smoke$Race, smoke$everSmoke, exclude=NULL)
```

```
##
##             yes   no <NA>
##   white    1328 8352  213
##   black     299 2977  154
##   hispanic 1054 4716  311
##   asian      93  858   22
##   native     35  291   12
##   pacific    23   56    6
##   <NA>      148  904  155
```

```
smokeSub = smoke[smoke$Age >= 10 & !is.na(smoke$Race) &
                   !is.na(smoke$everSmoke) & !is.na(smoke$ever_tobacco_hookah_or_wa) &!is.na(smoke$Sex), ]
dim(smokeSub)
```

```
## [1] 19986   163
```

```
smokeAgg = reshape2::dcast(smokeSub,
                           Age + Sex + Race + RuralUrban ~ everSmoke,
                           length)
```

```
## Using everSmoke as value column: use value.var to override.
```

```
dim(smokeAgg)
```

```
## [1] 208   7
```

```
smokeAgg = na.omit(smokeAgg)
dim(smokeAgg)
```

```
## [1] 207   7
```

```
smokeAgg$y = cbind(smokeAgg$yes, smokeAgg$no)
smokeFit = glm(y ~ Age + Sex + Race + RuralUrban,
               family=binomial(link='logit'), data=smokeAgg)

knitr::kable(summary(smokeFit)$coef, digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.003	0.186	-43.111	0.000
Age	0.419	0.012	36.266	0.000
SexF	0.042	0.043	0.980	0.327
Raceblack	-0.635	0.070	-9.005	0.000
Racehispanic	0.346	0.048	7.138	0.000
Raceasian	-0.631	0.118	-5.362	0.000
Racenative	0.160	0.190	0.838	0.402
Racepacific	0.964	0.270	3.566	0.000
RuralUrbanRural	-0.388	0.044	-8.769	0.000

```
smokeAgg$ageC = smokeAgg$Age - 15
smokeFit2 = glm(y ~ ageC + Sex + Race + RuralUrban,
                family=binomial(link='logit'), data=smokeAgg)
knitr::kable(summary(smokeFit2)$coef, digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.724	0.044	-39.226	0.000
ageC	0.419	0.012	36.266	0.000
SexF	0.042	0.043	0.980	0.327
Raceblack	-0.635	0.070	-9.005	0.000
Racehispanic	0.346	0.048	7.138	0.000
Raceasian	-0.631	0.118	-5.362	0.000
Racenative	0.160	0.190	0.838	0.402
Racepacific	0.964	0.270	3.566	0.000
RuralUrbanRural	-0.388	0.044	-8.769	0.000

```
#The odds of regular use of hookah
knitr::kable(exp(summary(smokeFit)$coef), digits=3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.000	1.204	0.000000e+00	1.000
Age	1.520	1.012	5.623341e+15	1.000

	Estimate	Std. Error	z value	Pr(>|z|)
SexF	1.043	1.044	2.666000e+00	1.387
Raceblack	0.530	1.073	0.000000e+00	1.000
Racehispanic	1.413	1.050	1.258299e+03	1.000
Raceasian	0.532	1.125	5.000000e-03	1.000
Racenative	1.173	1.210	2.312000e+00	1.495
Racepacific	2.621	1.310	3.538700e+01	1.000
RuralUrbanRural	0.678	1.045	0.000000e+00	1.000

```
# The CI of parameter
sum =summary(smokeFit2)
est=sum$coefficients[,1]
std=sum$coefficients[,2]

exp_upper_bound = exp(est+2*std)
exp_upper_bound
```

```
##     (Intercept)            ageC            SexF       Raceblack
##       0.1948306       1.5554041       1.1363689       0.6103322
##    Racehispanic       Raceasian      Racenative     Racepacific
##       1.5565615       0.6733007       1.7167054       4.4999523
## RuralUrbanRural
##       0.7409467
```

```
exp_lower_bound= exp(est-2*std)
exp_lower_bound
```

```
##     (Intercept)            ageC            SexF       Raceblack
##       0.1634291       1.4852154       0.9572099       0.4603920
##    Racehispanic       Raceasian      Racenative     Racepacific
##       1.2824617       0.4205311       0.8015719       1.5268970
## RuralUrbanRural
##       0.6206370
```

```
cbind(exp_lower_bound,exp_upper_bound)
```

```
##                 exp_lower_bound exp_upper_bound
## (Intercept)           0.1634291       0.1948306
## ageC                  1.4852154       1.5554041
## SexF                  0.9572099       1.1363689
## Raceblack             0.4603920       0.6103322
## Racehispanic          1.2824617       1.5565615
## Raceasian             0.4205311       0.6733007
## Racenative            0.8015719       1.7167054
## Racepacific           1.5268970       4.4999523
## RuralUrbanRural       0.6206370       0.7409467
```