Applied GLM Final Project

Lauren Bergam

April 21, 2019

```r
nacc <- read.csv("C:/Users/lberg/Documents/Brown/Applied GLM/Final Project/nacc_data_cognitive_clean.csv
```

```r
# Clean data, errant value for apoe_01
nacc <- nacc[(nacc$apoe_01 == 0 | nacc$apoe_01 == 1), ]

# Contingency table apoe and dementia diagnosis (1st term in function is exposure, 2nd is outcome)
table_apoe <- table(Exposure = factor(nacc$apoe_01, levels = c("1", "0")),
                    Outcome = factor(nacc$DEMENTED, levels = c("1", "0")))
table_apoe
```

```
##          Outcome
## Exposure     1     0
##        1   261  8876
##        0   232 20101
```

```r
# Fisher's test
fisher.test(table_apoe)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table_apoe
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   2.122623 3.059469
## sample estimates:
## odds ratio
##   2.547646
```

Interpretation: The odds of having dementia is 2.55 times more likely for individuals who are a carrier of the apolipoprotein E than those who are non-carriers. The chi-squared test gives a very small p-value and therefore, dementia and apoe carriers are not independent from each other.

```r
# Contingency table for age
nacc["age_threshold"] = ifelse(nacc$age_visit > 70, 1, 0)
age_table <- table("Age > 70" = factor(nacc$age_threshold, levels = c("1", "0")),
                   "Dementia" =  factor(nacc$DEMENTED, levels = c("1", "0")))
age_table
```

```
##          Dementia
## Age > 70     1     0
##        1   476 18515
##        0    17 10462
```

```
# Fisher's test
fisher.test(age_table)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  age_table
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   9.77067 27.41693
## sample estimates:
## odds ratio
##   15.82099
```

Since age is a discrete, non-binary variable, we choose a split value (in this case 70) to classify the data into two groups for a contingency table. Interpretation: The odds of having dementia is 15.8 times more likely for individuals over the age of 70 than those who are equal to or under the age of 70. The chi-squared test gives a very small p-value and therefore, dementia and age are not independent from each other.

```
# Contingency table for animal test
nacc["animal_threshold"] = ifelse(nacc$animal < 10, 1, 0)
animal_table <- table("Animal < 10 " = factor(nacc$animal_threshold, levels = c("1", "0")),
                      "Dementia" =  factor(nacc$DEMENTED, levels = c("1", "0")))
animal_table
```

```
##              Dementia
## Animal < 10      1     0
##            1   147   516
##            0   346 28461
```

```
# Fisher's Test
fisher.test(animal_table)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  animal_table
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   18.82025 29.09162
## sample estimates:
## odds ratio
##   23.43853
```

Since number of animals recalled is a discrete, non-binary variable, we choose a split value (in this case 10) to classify the data into two groups. Interpretation: The odds of having dementia is 23.4 times more likely for individuals who can recall fewer than 10 animals within one minute than those who can recall 10 or more. The chi-squared test gives a very small p-value and therefore, dementia and the results from the animal test are not independent from each other.