potassium

Stephen Shannon

June 30, 2021

```
library(dplyr)
library(tidyverse)
library(haven)
```

```
master <- read_dta("master_data.dta")
adverse <- read_dta("t05xc.dta")
lab <- read_dta("t016.dta")
```

General question: In this randomized study, what is the effect of the drug on elevated potassium levels (i.e. hyperkalemia)?

Questions: 1. Is there evidence that the drug is associated with hyperkalemia? 2. Does the drug effect depend on geographic region?

```
  #check for unique ids
x1 <- unique(lab$master_id)
x2 <- unique(adverse$master_id)
x3 <- unique(master$master_id)
#length(unique(c(x1,x2))) #3437
#length(unique(x3)) #3445

  #there is a difference of 8 ids in the master dataset that arent in the adverse
  #or lab dataset, they must be removed from the master dataset
x4 <- setdiff(x3, c(x1,x2)) #there are 9 values in x3 that are not in c(x1,x2)
x5 <- setdiff(c(x1,x2), x3) #there is 1 value in c(x1,x2) that is not in x3
master <- master %>% filter(!master_id %in% x4) #remove the patients with no
  #recorded k vals
  #extra entry in lab and adverse will be removed after the join of lab and
  #adverse

  #Construct new dataset for logistic regression
  #want k values from adverse, lab, then join together
  #adverse
  #filter all measurements that arent potassium related
  #then filter master_id and k
adverse <- adverse %>% filter(grepl(c('potassium'), ignore.case = TRUE, test_type)) %>% select(master_i

  #lab
  #repalce NA K values with 0, as they are assumed to be non hyper kalemic.
  #We are looking at a hard threshold of above or below 5.5, any value below
  #5.5 works to indicate non hyperkalemic. Then filter every col except master_id
```

```r
  #and k
lab <- lab %>% select(master_id, k) %>% mutate(k = replace_na(k, 0))

  #master
  #want treatment(1/0), region(1/2), and master_id
master <- master %>% select(master_id, treat, region)
  #make region a factor
master$region <- as.factor(master$region)

  #merge datasets adverse and lab together
df <- rbind(lab, adverse)
  #42,162 obs which is 725 + 41,437, successful merge
  #if a measurement is above 5.5, then for that measurement, indicate
  #hyperkalemia is present with a 1
df <- df %>% group_by(master_id) %>% mutate(hyperkalemic = ifelse(k >= 5.5, 1, 0))
  #if 1 hyperkalemic measurement is present for a particular id, change all to be
  #measurements to 1
df <- df %>% group_by(master_id) %>% mutate(hyperkalemic = ifelse(1 %in% hyperkalemic, 1, 0))
  #we do not need repeated measurements, so just  one obs per patient indicating
  #whether they were hyperkalemic at any point during the study
df <- df %>% group_by(master_id) %>% slice(1) %>% select(master_id, hyperkalemic)
df <- df %>% filter(!master_id %in% x5)
df <- merge(df, master, by="master_id")
#this will be the dataframe for doing logistic regression
```

```r
m1 <- glm(hyperkalemic ~ treat + region, family = "binomial", data=df)
summary(m1)
```

```
##
## Call:
## glm(formula = hyperkalemic ~ treat + region, family = "binomial",
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7109  -0.5592  -0.4835  -0.3754   2.3185
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.08756    0.09335 -22.363  < 2e-16 ***
## treat        0.84084    0.10524   7.990 1.35e-15 ***
## region2     -0.52978    0.10280  -5.153 2.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2749.9  on 3435  degrees of freedom
## Residual deviance: 2655.4  on 3433  degrees of freedom
## AIC: 2661.4
##
## Number of Fisher Scoring iterations: 5
```

```r
  #treat and region are signifigant, region 2 is associated with a decrease in the
  #log-odds of hyperkalemia (lower rate of hyperkalemia) compared to region 1,
  #while the drug is associated with an increase in the log-odds of hyperkalemia
  #compared to the placebo

m2 <- glm(hyperkalemic ~ treat*region, family = "binomial", data=df)
summary(m2)
```

```
##
## Call:
## glm(formula = hyperkalemic ~ treat * region, family = "binomial",
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7550  -0.5024  -0.4445  -0.4226   2.2182
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.3709     0.1207 -19.636  < 2e-16 ***
## treat           1.2617     0.1437   8.780  < 2e-16 ***
## region2         0.1057     0.1690   0.625    0.532
## treat:region2  -1.0026     0.2147  -4.670 3.01e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2749.9  on 3435  degrees of freedom
## Residual deviance: 2633.4  on 3432  degrees of freedom
## AIC: 2641.4
##
## Number of Fisher Scoring iterations: 5
```

```r
  #according to this model, the interaction between the drug and region is
  #signifigant. While the treatment has a positive log-odds for hyperkalemia
  #those who take the drug in region 2 still have higher log-odds than those
  #who dont take the drug, but have lower log-odds than those who take the drug
  #in region 1.

  #From these two models, the data suggests that there is an association between
  #the drug and hyperkalemia, and that the strength of this effect is depedent on
  #the geographic region
```