Data Analysis Project

BIOST 536

In order to make the household density variable that we are interested in in order to perform a grouped linear adjustment with ordinal variables, we must mutate our data by dividing family size by number of rooms.

```
tbdata = tbdata %>% mutate(famSize = case_when(
  famSize == "<4" ~ 1,
  famSize == "4-6" ~ 2,
  famSize == ">6" ~ 3
))


tbdata = tbdata %>% mutate(roomNo = case_when(
  roomNo == "1" ~ 1,
  roomNo == "2" ~ 2,
  roomNo == "3+" ~ 3
))

tbdata = tbdata %>% mutate(householdDensity = (famSize / roomNo))
```

Now, we can fit a logistic regression model to estimate the risk of TB infection associated with the number of windows in a household, adjusting for sex and our household density variable.

```
tb_mod_glm = glm(case ~ windowNo + male + householdDensity,
                 family=binomial,
                 data=tbdata)

summary(tb_mod_glm)

##
## Call:
## glm(formula = case ~ windowNo + male + householdDensity, family = binomial,
##     data = tbdata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6273  -1.1060  -0.0844   1.2330   1.4311
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.7711     0.2902    2.66   0.0079 **
## windowNo1         -1.0227     0.2529   -4.04  5.3e-05 ***
## windowNo2-3       -1.3772     0.2783   -4.95  7.5e-07 ***
## windowNo4+        -0.7026     0.3795   -1.85   0.0641 .
## male               0.0357     0.1864    0.19   0.8483
## householdDensity   0.0812     0.1430    0.57   0.5703
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 693.15  on 499  degrees of freedom
## Residual deviance: 662.68  on 494  degrees of freedom
## AIC: 674.7
##
## Number of Fisher Scoring iterations: 4
```

```
tb_mod_glm_null = glm(case ~ male + householdDensity,
                      family=binomial,
                      data=tbdata)
```

In this case, the model estimates that the odds of TB infection for a subpopulation of individuals with 1 window in their house, with the same sex and household density, is 0.36 times the odds of TB infection for a subpopulation of individuals without windows in their house with a 95% CI using model based SEs of [0.217, 0.586].

Comparing individuals with 2-3 windows vs no windows, the OR is now 0.252 with a 95% CI using model based SEs of [0.145, 0.432].

Comparing individuals with 4 or more windows vs no windows, the OR is now 0.495 with a 95% CI using model based SEs of [0.235, 1.044].

Since we have full and null models at this point, we can also run an LRtest in order to assess the significance of the number of windows in a house being a risk factor for TB infection. We'll use the significance level $\alpha = 0.05$

```
lrtest(tb_mod_glm, tb_mod_glm_null)
```

```
## Likelihood ratio test
##
## Model 1: case ~ windowNo + male + householdDensity
## Model 2: case ~ male + householdDensity
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1   6   -331
## 2   3   -346 -3  28.5    2.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```