

Data Analysis Project

Group 1

```
# mutate data to perform grouped linear adjustment with ordinal variables
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
))

# divide ordinal familySize by roomNo to create crude householdDensity covariate
tbdata = tbdata %>% mutate(householdDensity = (famSize / roomNo))

# fit logreg model to estimate risk of TB associated with windows, adjusting for sex, household density
tb_mod_glm = glm(case ~ windowNo + male + householdDensity + income,
  family=binomial,
  data=tbdata)

summary(tb_mod_glm)
```

```
##
## Call:
## glm(formula = case ~ windowNo + male + householdDensity + income,
##      family = binomial, data = tbdata)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.5578     0.3147   1.77  0.07637 .
## windowNo1     -0.9966     0.2574  -3.87  0.00011 ***
## windowNo2-3   -1.3068     0.2863  -4.57   5e-06 ***
## windowNo4+    -0.5615     0.3923  -1.43  0.15239
## male           0.0216     0.1890   0.11  0.90920
## householdDensity 0.0336     0.1462   0.23  0.81847
## incomelow      0.8798     0.2511   3.50  0.00046 ***
## incomemid      0.1667     0.2265   0.74  0.46173
## ---
## 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: 649.72 on 492 degrees of freedom
## AIC: 665.7
##
## Number of Fisher Scoring iterations: 4
```

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

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 0.369 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 [0.221, 0.607].

Comparing individuals with 2-3 windows vs no windows, the OR is now 0.271 with a 95% CI using model based SEs of [0.153, 0.471].

Comparing individuals with 4 or more windows vs no windows, the OR is now 0.57 with a 95% CI using model based SEs of [0.264, 1.233].

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 $\alpha = 0.05$

```
# perform LR test on full and null models to assess significance of numWindows
lrtest(tb_mod_glm, tb_mod_glm_null)
```

```
## Likelihood ratio test
##
## Model 1: case ~ windowNo + male + householdDensity + income
## Model 2: case ~ male + householdDensity + income
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 8 -325
## 2 5 -338 -3 25.4 1.3e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At the aforementioned significance level, we can reject the null hypothesis that the number of windows in a house does not have an effect on the risk of TB infection for this population of individuals.

```
# Set environment and load data
rm(list=ls())
library(tidyverse, warn.conflicts = F)
library(gt)
library(gtable)
library(gtsummary)

setwd('/Users/calenmendall/Desktop/UW_school_projects/Biost536/BIOST536_Extended_abstract/')
load(file = 'tbddata.RData')

## Add in household density measure
tbddata = tbddata %>% 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 = round((famSize / roomNo),2))

## Select relevant subset of data to table
tbdata_sub = tbdata %>% select(
  'Case' = case,
  'Number of Windows in home' = windowNo,
  'Sex' = male,
  'Age' = age,
  'Income' = income,
  'Family Size' = famSize,
  'Number of Rooms' = roomNo,
  'Household density' = householdDensity,
  'BCG Scar' = BCGscar,
  'Household member with TB' = houseMemberTb
)

## Some transformations for the table
tbdata_sub$Sex = factor(tbdata_sub$Sex,
  labels = c('Female', 'Male'))
tbdata_sub$Case = factor(tbdata_sub$Case,
  labels = c('Control', 'Case'))
tbdata_sub$Income = factor(tbdata_sub$Income,
  labels = c('Low', 'Mid', 'High'),
  levels = c('low', 'mid', 'high'))
tbdata_sub$`Number of Rooms` = factor(tbdata_sub$`Number of Rooms`,
  labels = c('1', '2', '3+'),
  levels = c(1, 2, 3))
tbdata_sub$`Family Size` = factor(tbdata_sub$`Family Size`,
  labels = c('<4', '4-6', '>6'),
  levels = c(1, 2, 3))
tbdata_sub$`BCG Scar` = factor(tbdata_sub$`BCG Scar`,
  labels = c('No', 'Yes'))
tbdata_sub$`Household member with TB` = factor(tbdata_sub$`Household member with TB`,
  labels = c('No', 'Yes'))

## Generating the table and formatting
tab = tbl_summary(tbdata_sub, by = 'Case') %>% add_overall(last=T)%>% as_gt()

tab = tab %>% tab_spanner('Tuberculosis infection status', columns = c(6,7,8), gather = T, level = 1)

tab2 = tab %>% tab_options(table.font.size = 14, heading.title.font.size = 15,
  row_group.font.weight = 'bold',
  column_labels.font.weight = 'bold',

```

```

        row_group.background.color = "gray95",
        column_labels.background.color = 'gray80', stub.border.width = 0,
        data_row.padding.horizontal = 0,
        data_row.padding = 2)

tab2 = tab2 %>% cols_width(everything()~185)
tab2
## Saving the table

gtsave(tab2, filename = "table1.pdf", path = "./")
gtExtras::gtsave_extra(tab2, filename = "table1.2.png",path = "./")

```

```

flowchart LR
    A{Number of \nwindows\nin home} --> B{Tuberculosis\ninfection}
    C{Sex} --> B
    A --> D{Household members \nwith tuberculosis}
    D --> B
    H{BCG\nvaccination} --> B
    I{Smoking} --> B
    E{Household density} --> B
    E --> A
    F{Income}--> E
    G{Age} --> F
    C --> F
    G --> B
    F --> A

```