

Non Parametric Model

Christine Long

April 19, 2019

```
#Load Libraries
library(tidyverse)
```

```
## -- Attaching packages -----
----- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  1.4.2      v dplyr   0.7.8
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0
```

```
## -- Conflicts -----
- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(gee)
```

```
## Warning: package 'gee' was built under R version 3.5.3
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
##
##      expand
```

```
#Load data
health = readxl::read_excel("HW8-HEALTH.xlsx") %>%
  janitor::clean_names()
```

- a. Evaluate the bivariate, cross-sectional relationship between randomized group assignment and participants health self-rating at the time of randomization.

```

bas_health = health[health$time==1,]

#Proportional table
#Control Group at baseline
ctrl = table(bas_health$txt,bas_health$health)[1,]/sum(table(bas_health$txt,bas_health$health)[1,])
ctrl

```

```

##           Good           Poor
## 0.4878049 0.5121951

```

```

#Treatment Group at baseline
trt = table(bas_health$txt,bas_health$health)[2,]/sum(table(bas_health$txt,bas_health$health)[2,])
trt

```

```

##           Good           Poor
## 0.4102564 0.5897436

```

```

rbind(ctrl,trt)

```

```

##           Good           Poor
## ctrl 0.4878049 0.5121951
## trt   0.4102564 0.5897436

```

```

prop.test(rbind(ctrl,trt))

```

```

## Warning in prop.test(rbind(ctrl, trt)): Chi-squared approximation may be
## incorrect

```

```

##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  rbind(ctrl, trt)
## X-squared = 1.1186e-32, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -1  1
## sample estimates:
##   prop 1    prop 2
## 0.4878049 0.4102564

```

As we can see from the proportional group, there is basically no difference between control and treatment group at baseline / at the time of randomization. By using two sample proportional test, p value is almost 1, meaning there is not difference between the control and treatment groups in terms of health ratings.

This make sense, because the treatment hasn't started yet, and given well randomization, the baseline characters for the two groups should be approximately same in differenet treatment groups.

(b) Perform a longitudinal data analysis across all study follow-up visits (but not at randomization) to describe the relationship of the participants self-ratings as a function of the effects of health self-rating at the baseline, treatment group, month post randomization, and age group as predictors. Fit a GEE with unstructured correlation structure. Interpret your results.

```
fo_health = health[health$time != 1,]
full_health = bas_health[,c(1,4)] %>%
  right_join(fo_health, by = "id") %>%
  rename(bashealth = health.x, fohealth = health.y) %>%
  mutate(time = recode(time, '2' = '3', '3' = '6', '4' = '12'),
         time = as.numeric(time),
         fohealth = as.numeric(fohealth == "Good"))

gee.fit = gee(fohealth~bashealth+time+txt+agegroup, data = full_health, family = "binomial", id
= id,
              constr = "unstructured", scale.fix = TRUE, scale.value = 1)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
##      (Intercept)  bashealthPoor      time txtIntervention
##      0.18528086   -1.71063852    0.02536275    1.99669985
##   agegroup25-34   agegroup35+
##      1.19749448    1.39742621
```

```
summary(gee.fit)
```

```
##
## GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
## gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
## Link:                               Logit
## Variance to Mean Relation: Binomial
## Correlation Structure:               Unstructured
##
## Call:
## gee(formula = fohealth ~ bashealth + time + txt + agegroup, id = id,
##      data = full_health, family = "binomial", constr = "unstructured",
##      scale.fix = TRUE, scale.value = 1)
##
## Summary of Residuals:
##           Min           1Q           Median           3Q           Max
## -0.98144969 -0.18317233  0.08914345  0.17159228  0.83093959
##
##
## Coefficients:
##              Estimate Naive S.E.   Naive z Robust S.E.   Robust z
## (Intercept)    0.12457924 0.47137316  0.2642901  0.51374172  0.2424939
## bashealthPoor -1.81418056 0.48958528 -3.7055456  0.50961334 -3.5599158
## time           0.03243343 0.03665686  0.8847848  0.04755408  0.6820326
## txtIntervention 2.10225898 0.48779381  4.3097286  0.53777951  3.9091467
## agegroup25-34  1.35250468 0.48130172  2.8100973  0.50420159  2.6824681
## agegroup35+    1.42052166 0.79781620  1.7805124  0.78372968  1.8125148
##
## Estimated Scale Parameter:  1
## Number of Iterations:  5
##
## Working Correlation
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.1719328 0.5859907
## [2,] 0.1719328 1.0000000 0.2013998
## [3,] 0.5859907 0.2013998 1.0000000
```

Interpretation:

- Intercept: For patients with Good baseline health rating and age within 15-24 at control group at baseline, the log odds of good health rating is 0.12457924 on average
- bashealthPoor: For patients within the same age group and same treatment group at same post randomization month, the log odds ratio of good health rating for poor baseline rating vs. good baseline rating is -1.81418056 on average
- Time: For patients within the same age group and same treatment group and same baseline rating, the log odds ratio of good health rating for 1 month increase is 0.03243343 on average
- txtIntervention: For patients within the same age group and same baseline rating at same post randomization month, the log odds ratio of good health rating for 1 month increase is 2.10225898 on average
- agegroup25-34: For patients within the same treatment group and same baseline rating at same post randomization month, the log odds ratio of good health rating for age within 25-34 vs. age within 15-24 is 1.35250468 on average

- agegroup35+: For patients within the same treatment group and same baseline rating at same post randomization month, the log odds ratio of good health rating for age above 35 vs. age within 15-24 is 1.42052166 on average

(c) Fit a generalized linear mixed effects model with subject-specific random intercepts.

```
glmm.fit = glmer(fohealth~bashealth+time+txt+agegroup + (1 | id),data = full_health, family = "binomial")
summary(glmm.fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: fohealth ~ bashealth + time + txt + agegroup + (1 | id)
## Data: full_health
##
##      AIC      BIC    logLik deviance df.resid
##    185.0    208.0    -85.5    171.0      192
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6112 -0.2327  0.1402  0.2982  1.8239
##
## Random effects:
## Groups Name      Variance Std.Dev.
## id      (Intercept) 5.721    2.392
## Number of obs: 199, groups: id, 78
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.19521    0.87019   0.224  0.82250
## bashealthPoor -2.77610    0.98381 -2.822  0.00478 **
## time           0.03718    0.06933   0.536  0.59176
## txtIntervention 3.41325    1.07268   3.182  0.00146 **
## agegroup25-34  2.25651    1.00877   2.237  0.02529 *
## agegroup35+    1.98229    1.38118   1.435  0.15123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) bshltP time   txtInt a25-34
## bashealthPr -0.374
## time        -0.472 -0.016
## txtIntrvntn -0.256 -0.449  0.047
## agegrp25-34 -0.319 -0.379  0.007  0.395
## agegroup35+ -0.195 -0.274 -0.007  0.206  0.390
```

Interpretation:

- Intercept: For the same patients with Good baseline health rating and age within 15-24 at control group at baseline, the log odds of good health rating is 0.19521
- Time: For patients within the same age group and same treatment group and same baseline rating, the log odds ratio of good health rating for one month increase is 0.03718

- `bashealthPoor`, `agegroup25-34`, `agegroup35+` cannot be interpreted, because for the same subject, these characteristics are fixed, and there is no point in comparing it to different levels.

Difference between `glmm` and `gee`: `glmm` focuses on individual level, therefore `bashealthPoor`, `agegroup25-34`, `agegroup35+` cannot be interpreted; Meanwhile, `gee` focuses on sub-population level, and the interpretation is the average over the sub-population.