Day 1: Generalized Linear Mixed Models

Back to schedule

Read in the individual data (or a pairwise dataset)

```r
library(tidyr)
library(dplyr)

#install.packages("lme4")
library(lme4)

acitelli_ind <- read.csv(file.choose(), header=TRUE)
```

Convert individual data to pairwise. I also create a simhobs variable that will be our binary response, two dummy variables that will be useful for estimating separate random intercepts for men and women, and a count variable `cigarettes`.

```r
tempA <- acitelli_ind %>%
  mutate(genderE = gender, partnum = 1) %>%
  mutate(gender = ifelse(gender == 1, "A", "P")) %>%
  gather(variable, value, self_pos:genderE) %>%
  unite(var_gender, variable, gender) %>%
  spread(var_gender, value)

tempB <- acitelli_ind %>%
  mutate(genderE = gender, partnum = 2) %>%
  mutate(gender = ifelse(gender == 1, "P", "A")) %>%
  gather(variable, value, self_pos:genderE)%>%
  unite(var_gender, variable, gender) %>%
  spread(var_gender, value)

acitelli_pair <- bind_rows(tempA, tempB) %>%
  arrange(cuplid) %>%
  mutate(gender_A = ifelse(genderE_A == 1, "hus", "wife"),
         gender_A = as.factor(gender_A),
         simhob_bin_A = ifelse(simhob_A == 1, 1, 0), #forced binary variable
         man = ifelse(genderE_A == 1, 1, 0),
         woman = ifelse(genderE_A == 1, 0, 1),
         cigarettes_A = rpois(296, 0.7)) %>%
  group_by(cuplid) %>%
  mutate(cupcig = rpois(1, 0.7)) %>%
  ungroup(cuplid) %>%
  mutate(cigarettes_A = cigarettes_A + cupcig)

rm(tempA, tempB)
```

Logistic Regression and Poisson Regression

First, let's only select the women so we can do away with the nesting.

```
acitelli_women <- acitelli_pair %>%
  filter(genderE_A == -1)
```

Logistic Regression

Then we run a ordinary logistic regression.

```
logistic_reg <- glm(simhob_bin_A ~ other_pos_A + other_pos_P,
                    data = acitelli_women,
                    family = binomial)
```

```
summary(logistic_reg)
```

```
##
## Call:
## glm(formula = simhob_bin_A ~ other_pos_A + other_pos_P, family = binomial,
##     data = acitelli_women)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1161  -0.8659  -0.6774   1.2562   1.9842
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.7461     2.1892  -2.625  0.00867 **
## other_pos_A   0.9704     0.4032   2.407  0.01609 *
## other_pos_P   0.1496     0.4081   0.367  0.71387
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 176.56  on 147  degrees of freedom
## Residual deviance: 169.21  on 145  degrees of freedom
## AIC: 175.21
##
## Number of Fisher Scoring iterations: 4
```

This gives us the increase in log odds of a woman having the same hobbies as her husband for every 1 unit increase in other positivity of the actor and partner other positivity.

If we want to easily switch estimates from "log odds" to "odds," we use the following:

```
exp(logistic_reg$coefficients)
```

```
## (Intercept) other_pos_A other_pos_P
## 0.003195341 2.639004851 1.161424308
```

Poisson Regression (aka, Log-Linear Regression)

First, we run the model assuming the variance equals the mean.

```
poisson_reg <- glm(cigarettes_A ~ other_pos_A + other_pos_P,
                   data = acitelli_women,
```

```
                   family = poisson)

summary(poisson_reg)

##
## Call:
## glm(formula = cigarettes_A ~ other_pos_A + other_pos_P, family = poisson,
##     data = acitelli_women)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7747  -0.4500  -0.3417   0.5048   1.8388
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.80970    0.75073   1.079    0.281
## other_pos_A -0.03731    0.13464  -0.277    0.782
## other_pos_P -0.06728    0.14826  -0.454    0.650
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 157.32  on 147  degrees of freedom
## Residual deviance: 156.96  on 145  degrees of freedom
## AIC: 444.23
##
## Number of Fisher Scoring iterations: 5

exp(poisson_reg$coefficients)

## (Intercept) other_pos_A other_pos_P
##   2.2472293   0.9633735   0.9349347
```

But, we can also run a model relaxing the variance = mean assumption. That is, we can include an over-dispersion parameter with `family = quasipoisson`.

```
poissonOD_reg <- glm(cigarettes_A ~ other_pos_A + other_pos_P,
                     data = acitelli_women,
                     family = quasipoisson)

summary(poissonOD_reg)

##
## Call:
## glm(formula = cigarettes_A ~ other_pos_A + other_pos_P, family = quasipoisson,
##     data = acitelli_women)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7747  -0.4500  -0.3417   0.5048   1.8388
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.80970    0.70154   1.154    0.250
## other_pos_A -0.03731    0.12582  -0.297    0.767
## other_pos_P -0.06728    0.13854  -0.486    0.628
##
```

```
## (Dispersion parameter for quasipoisson family taken to be 0.8732518)
##
##     Null deviance: 157.32  on 147  degrees of freedom
## Residual deviance: 156.96  on 145  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

```r
exp(poissonOD_reg$coefficients)
```

```
## (Intercept) other_pos_A other_pos_P
##    2.2472293   0.9633735   0.9349347
```

Logistic Multilevel Modeling (Binary variables)

To account for the nonindependence, we can make use of the `glmer()` function from the `lme4` package. Note that we are asking for the variance of intercepts across dyads, that is the random intercept in traditonal multilevel modeling. The `gls()` function in the `nlme` package does not have an option for specifying a link function (i.e., there is no `family =` option). The syntax of `glmer()` differs a bit from `gls()` in that the random effects are specified within the formula: `+ (1/cuplid)`.

Indistinguishable Dyads

```r
apim_bin <- glmer(simhob_bin_A ~ other_pos_A + other_pos_P
                  + (1|cuplid),
                  data = acitelli_pair,
                  family = binomial,
                  na.action = na.omit)
```

```r
summary(apim_bin)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: simhob_bin_A ~ other_pos_A + other_pos_P + (1 | cuplid)
##    Data: acitelli_pair
##
##      AIC      BIC   logLik deviance df.resid
##    322.5    337.3   -157.3    314.5      292
##
## Scaled residuals:
##     Min      1Q   Median       3Q      Max
## -0.90316 -0.44848 -0.33245  0.02569  2.03565
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  cuplid (Intercept) 1.659    1.288
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -8.3310     2.4903  -3.345 0.000821 ***
```

```
## other_pos_A    0.9195      0.3894    2.361 0.018217 *
## other_pos_P    0.6723      0.3756    1.790 0.073481 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) oth__A
## other_pos_A -0.738
## other_pos_P -0.708  0.058
```

Distinguishable Dyads

Interaction approach.

```r
apim_bin_di <- glmer(simhob_bin_A ~ other_pos_A + other_pos_P + genderE_A
                     + other_pos_A*genderE_A + other_pos_P*genderE_A
                     + (man + woman - 1|cuplid),
                     data = acitelli_pair,
                     family = binomial,
                     na.action = na.omit)

summary(apim_bin_di)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## simhob_bin_A ~ other_pos_A + other_pos_P + genderE_A + other_pos_A *
##     genderE_A + other_pos_P * genderE_A + (man + woman - 1 |     cuplid)
##    Data: acitelli_pair
##
##      AIC      BIC   logLik deviance df.resid
##    314.7    347.9   -148.4    296.7      287
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -1.09002 -0.48703 -0.00949  0.03292  2.52080
##
## Random effects:
##  Groups Name  Variance Std.Dev. Corr
##  cuplid man    183.878  13.560
##         woman    0.191   0.437  1.00
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -3.842e+01  6.918e-04  -55536   <2e-16 ***
## other_pos_A          6.104e+00  6.917e-04    8824   <2e-16 ***
## other_pos_P          1.435e+00  6.917e-04    2074   <2e-16 ***
## genderE_A           -3.062e+01  6.918e-04  -44265   <2e-16 ***
## other_pos_A:genderE_A 5.015e+00  6.917e-04    7250   <2e-16 ***
## other_pos_P:genderE_A 9.391e-01  6.917e-04    1358   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##            (Intr) oth__A oth__P gndE_A o__A:E
## other_pos_A 0.000
## other_pos_P 0.000  0.000
## genderE_A   0.000  0.000  0.000
## othr__A:E_A 0.000  0.000  0.000  0.000
## othr__P:E_A 0.000  0.000  0.000  0.000  0.000
## convergence code: 0
## Model failed to converge with max|grad| = 0.103721 (tol = 0.001, component 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

Two-intercept model.

```
apim_bin_di_two <- glmer(simhob_bin_A ~ gender_A + other_pos_A:gender_A + other_pos_P:gender_A - 1
                       + (man + woman - 1|cuplid),
                       data = acitelli_pair,
                       family = binomial,
                       na.action = na.omit)

summary(apim_bin_di_two)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## simhob_bin_A ~ gender_A + other_pos_A:gender_A + other_pos_P:gender_A -
##     1 + (man + woman - 1 | cuplid)
##    Data: acitelli_pair
##
##     AIC      BIC   logLik deviance df.resid
##   252.6    285.8   -117.3    234.6      287
##
## Scaled residuals:
##      Min        1Q    Median        3Q       Max
## -0.013522 -0.004989 -0.000136  0.002841  0.134981
##
## Random effects:
##  Groups Name  Variance  Std.Dev. Corr
##  cuplid man   172856.6  415.76
##         woman    860.6   29.34   0.02
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|)
## gender_Ahus             -27.8603    59.6381  -0.467    0.640
## gender_Awife            -19.3113    18.2034  -1.061    0.289
## gender_Ahus:other_pos_A   1.8910     7.4050   0.255    0.798
## gender_Awife:other_pos_A  1.5086     1.8566   0.813    0.416
## gender_Ahus:other_pos_P   0.2470     2.8218   0.088    0.930
## gender_Awife:other_pos_P  0.6637     2.7059   0.245    0.806
##
## Correlation of Fixed Effects:
##             gndr_Ah gndr_Aw gndr_Ah:__A gndr_Aw:__A gndr_Ah:__P
```

```
## gender_Awif  0.810
## gndr_Ah:__A -0.932  -0.726
## gndr_Aw:__A -0.028  -0.375  -0.003
## gndr_Ah:__P -0.332  -0.268   0.029        0.150
## gndr_Aw:__P -0.737  -0.847   0.681       -0.122        0.179
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
## failure to converge in 10000 evaluations
```

Log-Linear Multilevel Modeling (Count variables)

Indistinguishable Dyads

```
apim_poi <- glmer(cigarettes_A ~ other_pos_A + other_pos_P
                  + (1|cuplid),
                  data = acitelli_pair,
                  family = poisson,
                  na.action = na.omit)

summary(apim_poi)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: cigarettes_A ~ other_pos_A + other_pos_P + (1 | cuplid)
##    Data: acitelli_pair
##
##      AIC      BIC   logLik deviance df.resid
##    861.4    876.1   -426.7    853.4      292
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2471 -0.4031 -0.2149  0.5547  2.1955
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  cuplid (Intercept) 0.06173  0.2485
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.2969     0.5717   2.269   0.0233 *
## other_pos_A  -0.1242     0.1024  -1.212   0.2255
## other_pos_P  -0.1073     0.1026  -1.045   0.2960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) oth__A
## other_pos_A -0.650
## other_pos_P -0.652 -0.144
```

```
## convergence code: 0
## Model failed to converge with max|grad| = 0.00332975 (tol = 0.001, component 1)
```

Distinguishable Dyads

Interaction approach.

```
apim_poi_di <- glmer(cigarettes_A ~ other_pos_A + other_pos_P + genderE_A
                     + other_pos_A*genderE_A + other_pos_P*genderE_A
                     + (man + woman - 1|cuplid),
                     data = acitelli_pair,
                     family = poisson,
                     na.action = na.omit)

summary(apim_poi_di)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula:
## cigarettes_A ~ other_pos_A + other_pos_P + genderE_A + other_pos_A *
##     genderE_A + other_pos_P * genderE_A + (man + woman - 1 |      cuplid)
##    Data: acitelli_pair
##
##      AIC      BIC   logLik deviance df.resid
##    869.8    903.0   -425.9    851.8      287
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2025 -0.4018 -0.2113  0.5282  2.5458
##
## Random effects:
##  Groups Name  Variance Std.Dev. Corr
##  cuplid man   0.05229  0.2287
##         woman 0.07277  0.2698   1.00
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.31387    0.57669   2.278   0.0227 *
## other_pos_A         -0.12796    0.10346  -1.237   0.2162
## other_pos_P         -0.10770    0.10352  -1.040   0.2982
## genderE_A            0.57183    0.52769   1.084   0.2785
## other_pos_A:genderE_A -0.10585    0.10549  -1.003   0.3157
## other_pos_P:genderE_A -0.03268    0.10557  -0.310   0.7569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) oth__A oth__P gndE_A o__A:E
## other_pos_A -0.649
## other_pos_P -0.651 -0.146
## genderE_A   -0.008 -0.074  0.079
## othr__A:E_A -0.080  0.087  0.018 -0.584
```

```
## othr__P:E_A  0.083  0.001 -0.103 -0.586 -0.308
## convergence code: 0
## Model failed to converge with max|grad| = 0.0769463 (tol = 0.001, component 1)
```

Two-intercept model.

```
apim_poi_di_two <- glmer(cigarettes_A ~ gender_A + other_pos_A:gender_A + other_pos_P:gender_A - 1
                         + (man + woman - 1|cuplid),
                         data = acitelli_pair,
                         family = poisson,
                         na.action = na.omit)
```

```
summary(apim_poi_di_two)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: poisson  ( log )
## Formula:
## cigarettes_A ~ gender_A + other_pos_A:gender_A + other_pos_P:gender_A -
##     1 + (man + woman - 1 | cuplid)
##    Data: acitelli_pair
##
##      AIC      BIC   logLik deviance df.resid
##    869.8    903.0   -425.9    851.8      287
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2028 -0.3994 -0.2102  0.5315  2.5502
##
## Random effects:
##  Groups Name  Variance Std.Dev. Corr
##  cuplid man   0.05249  0.2291
##         woman 0.07296  0.2701   1.00
## Number of obs: 296, groups:  cuplid, 148
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## gender_Ahus               1.89692    0.77869   2.436   0.0148 *
## gender_Awife              0.75124    0.78485   0.957   0.3385
## gender_Ahus:other_pos_A  -0.23412    0.15412  -1.519   0.1287
## gender_Awife:other_pos_A -0.03607    0.14097  -0.256   0.7981
## gender_Ahus:other_pos_P  -0.14287    0.14004  -1.020   0.3076
## gender_Awife:other_pos_P -0.06345    0.15548  -0.408   0.6832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             gndr_Ah gndr_Aw gndr_Ah:__A gndr_Aw:__A gndr_Ah:__P
## gender_Awif  0.089
## gndr_Ah:__A -0.669  -0.059
## gndr_Aw:__A -0.049  -0.561  -0.020
## gndr_Ah:__P -0.570  -0.049  -0.222       0.085
## gndr_Aw:__P -0.058  -0.666   0.086      -0.237      -0.020
## convergence code: 0
## Model failed to converge with max|grad| = 0.235447 (tol = 0.001, component 1)
```

Generalized Estimating Equations (GEE)

Indistinguishable Dyads

```r
#install.packages("gee")
library(gee)

apim_gee <- gee(simhob_bin_A ~ other_pos_A + other_pos_P,
                id = cuplid,
                data = acitelli_pair,
                na.action = na.omit,
                family = binomial,
                corstr = "unstructured")
```

```
## (Intercept) other_pos_A other_pos_P
##  -6.2582893   0.6956734   0.5017679
```

```r
summary(apim_gee)
```

```
##
##  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 = simhob_bin_A ~ other_pos_A + other_pos_P, id = cuplid,
##     data = acitelli_pair, na.action = na.omit, family = binomial,
##     corstr = "unstructured")
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
## -0.43310561 -0.27624377 -0.19297150  0.07522507  0.86606888
##
##
## Coefficients:
##                Estimate Naive S.E.   Naive z Robust S.E.  Robust z
## (Intercept) -6.2729871  1.8140043 -3.458088   1.7415077 -3.602044
## other_pos_A  0.6982148  0.2901570  2.406334   0.2851282  2.448775
## other_pos_P  0.5025443  0.2840654  1.769115   0.2605414  1.928846
##
## Estimated Scale Parameter:  0.9978716
## Number of Iterations:  2
##
## Working Correlation
##           [,1]      [,2]
## [1,] 1.0000000 0.2703261
## [2,] 0.2703261 1.0000000
```

Back to schedule