Modeling_part_draft

Bulun Te

2024-04-15

Longitudinal Possion Regression with random effect

assuming 2307 is male and 5094 is female

```
# constructing variable
load("complete_data3_2307_1_5094_0.RData")

complete_data  = complete_data %>% mutate(
  Pickups_lag_log = log(Pickups_lag),
  Total.ST.min.log = log(Total.ST.min)
)

complete_data <- complete_data %>%
  group_by(pseudo_ID) %>%
  mutate(success_lag = lag(success, order_by = Date),
         Total.ST.min.lag = lag(Total.ST.min, order_by = Date),
         Proportion.ST_lag = lag(Proportion.ST, order_by = Date))


library(lme4)
```

```
##      Matrix
```

```
model_1 <-
  glmer(
    Pickups ~   Pickups_lag_log +  Treatment + Phase + Proportion.ST +sex + Semester:Xt + devices + sex
    data = complete_data,
    family = poisson(link = "log"),
    control = glmerControl(optCtrl = list(maxfun = 1e5),optimizer = "bobyqa")
  )

summary(model_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Pickups ~ Pickups_lag_log + Treatment + Phase + Proportion.ST +
##     sex + Semester:Xt + devices + sex:log(Total.ST.min.lag) +
```

```
##     (1 + Semester:Xt | pseudo_ID)
##    Data: complete_data
## Control: glmerControl(optCtrl = list(maxfun = 1e+05), optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  37618.1  37686.9 -18797.1  37594.1     2262
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.2587  -1.9878  -0.2166   1.6095  22.9736
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  pseudo_ID (Intercept) 0.17175  0.4144
##            Semester:Xt 0.02007  0.1417   -0.75
## Number of obs: 2274, groups:  pseudo_ID, 25
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                2.584714   0.211345  12.230  < 2e-16 ***
## Pickups_lag_log            0.318458   0.005818  54.735  < 2e-16 ***
## TreatmentB                -0.101124   0.141216  -0.716   0.4739
## Phase                     -0.297733   0.010826 -27.501  < 2e-16 ***
## Proportion.ST              0.183177   0.014985  12.224  < 2e-16 ***
## sex                       -0.195232   0.132776  -1.470   0.1415
## devices                    0.132773   0.062544   2.123   0.0338 *
## Semester:Xt                0.148790   0.028835   5.160 2.47e-07 ***
## sex:log(Total.ST.min.lag)  0.030771   0.007355   4.184 2.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Pckp__ TrtmnB Phase  Prp.ST sex    devics Smst:X
## Pckps_lg_lg -0.108
## TreatmentB  -0.130  0.013
## Phase       -0.013  0.137  0.007
## Proportn.ST -0.032  0.022  0.006 -0.053
## sex         -0.296  0.019 -0.106 -0.017 -0.023
## devices     -0.755 -0.023 -0.262 -0.006 -0.007 -0.029
## Semester:Xt -0.291 -0.022 -0.023 -0.011 -0.009  0.016  0.009
## sx:(T.ST..)  0.016 -0.063  0.015  0.050  0.066 -0.324 -0.014 -0.006
```

```
car::vif(model_1)
```

```
##           Pickups_lag_log                 Treatment                     Phase
##                  1.026542                  1.090258                  1.026741
##             Proportion.ST                       sex                   devices
##                  1.008906                  1.135215                  1.079175
##               Semester:Xt sex:log(Total.ST.min.lag)
##                  1.001360                  1.131497
```

assuming 2307 is male and 5094 is male

```
# constructing variable
load("complete_data3_2307_1_5094_1.RData")

complete_data  = complete_data %>% mutate(
  Pickups_lag_log = log(Pickups_lag),
  Total.ST.min.log = log(Total.ST.min)
)

complete_data <- complete_data %>%
  group_by(pseudo_ID) %>%
  mutate(success_lag = lag(success, order_by = Date),
         Total.ST.min.lag = lag(Total.ST.min, order_by = Date),
         Proportion.ST_lag = lag(Proportion.ST, order_by = Date))


library(lme4)

model_2 <-
  glmer(
    Pickups ~   Pickups_lag_log +  Treatment + Phase + Proportion.ST +sex + Semester:Xt + devices + sex
    data = complete_data,
    family = poisson(link = "log"),
    control = glmerControl(optCtrl = list(maxfun = 1e5),optimizer = "bobyqa")
  )

summary(model_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Pickups ~ Pickups_lag_log + Treatment + Phase + Proportion.ST +
##     sex + Semester:Xt + devices + sex:log(Total.ST.min.lag) +
##     (1 + Semester:Xt | pseudo_ID)
##    Data: complete_data
## Control: glmerControl(optCtrl = list(maxfun = 1e+05), optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  37628.0  37696.8 -18802.0  37604.0     2262
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -10.222  -1.988  -0.219   1.609  23.075
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  pseudo_ID (Intercept) 0.17746  0.4213
##            Semester:Xt 0.02022  0.1422   -0.75
## Number of obs: 2274, groups:  pseudo_ID, 25
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                2.572034   0.220059  11.688  < 2e-16 ***
## Pickups_lag_log            0.318935   0.005819  54.809  < 2e-16 ***
## TreatmentB                -0.106184   0.141680  -0.749  0.45358
## Phase                     -0.298100   0.010836 -27.511  < 2e-16 ***
## Proportion.ST              0.182076   0.014989  12.148  < 2e-16 ***
## sex                       -0.115885   0.132582  -0.874  0.38208
## devices                    0.133771   0.063060   2.121  0.03389 *
## Semester:Xt                0.149075   0.028939   5.151 2.59e-07 ***
## sex:log(Total.ST.min.lag)  0.019884   0.007175   2.771  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Pckp__ TrtmnB Phase  Prp.ST sex    devics Smst:X
## Pckps_lg_lg -0.103
## TreatmentB  -0.194  0.013
## Phase       -0.012  0.136  0.006
## Proportn.ST -0.033  0.021  0.006 -0.051
## sex         -0.369  0.020  0.079 -0.021 -0.018
## devices     -0.725 -0.022 -0.267 -0.006 -0.007 -0.040
## Semester:Xt -0.284 -0.022 -0.020 -0.011 -0.009  0.009  0.008
## sx:(T.ST..)  0.013 -0.066  0.013  0.065  0.074 -0.313 -0.013 -0.006
```

assuming 2307 is female and 5094 is female

```r
# constructing variable
load("complete_data3_2307_0_5094_0.RData")

complete_data  = complete_data %>% mutate(
  Pickups_lag_log = log(Pickups_lag),
  Total.ST.min.log = log(Total.ST.min)
)

complete_data <- complete_data %>%
  group_by(pseudo_ID) %>%
  mutate(success_lag = lag(success, order_by = Date),
         Total.ST.min.lag = lag(Total.ST.min, order_by = Date),
         Proportion.ST_lag = lag(Proportion.ST, order_by = Date))


library(lme4)

model_3 <-
  glmer(
    Pickups ~   Pickups_lag_log +  Treatment + Phase + Proportion.ST +sex + Semester:Xt + devices + sex
    data = complete_data,
    family = poisson(link = "log"),
    control = glmerControl(optCtrl = list(maxfun = 1e5),optimizer = "bobyqa")
  )

summary(model_3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Pickups ~ Pickups_lag_log + Treatment + Phase + Proportion.ST +
##     sex + Semester:Xt + devices + sex:log(Total.ST.min.lag) +
##     (1 + Semester:Xt | pseudo_ID)
##    Data: complete_data
## Control: glmerControl(optCtrl = list(maxfun = 1e+05), optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  37632.2  37701.0 -18804.1  37608.2     2262
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.2069  -1.9883  -0.2151   1.6260  23.2726
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  pseudo_ID (Intercept) 0.17703  0.4208
##            Semester:Xt 0.02027  0.1424   -0.75
## Number of obs: 2274, groups:  pseudo_ID, 25
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 2.580733   0.217534  11.864  < 2e-16 ***
## Pickups_lag_log             0.319144   0.005825  54.785  < 2e-16 ***
## TreatmentB                 -0.105896   0.141658  -0.748   0.4547
## Phase                      -0.298750   0.010836 -27.571  < 2e-16 ***
## Proportion.ST               0.180480   0.014967  12.059  < 2e-16 ***
## sex                        -0.103396   0.129746  -0.797   0.4255
## devices                     0.134575   0.063384   2.123   0.0337 *
## Semester:Xt                 0.149352   0.028974   5.155 2.54e-07 ***
## sex:log(Total.ST.min.lag)  0.014362   0.007753   1.852   0.0640 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Pckp__ TrtmnB Phase  Prp.ST sex    devics Smst:X
## Pckps_lg_lg -0.104
## TreatmentB  -0.160  0.012
## Phase       -0.013  0.135  0.007
## Proportn.ST -0.028  0.022  0.005 -0.053
## sex         -0.329  0.025 -0.014 -0.022 -0.028
## devices     -0.762 -0.022 -0.264 -0.006 -0.007  0.017
## Semester:Xt -0.287 -0.022 -0.021 -0.011 -0.009  0.012  0.009
## sx:(T.ST..)  0.012 -0.080  0.019  0.065  0.052 -0.344 -0.011 -0.003
```

assuming 2307 is female and 5094 is male

```
# constructing variable
load("complete_data3_2307_0_5094_1.RData")
```

```r
complete_data  = complete_data %>% mutate(
  Pickups_lag_log = log(Pickups_lag),
  Total.ST.min.log = log(Total.ST.min)
)

complete_data <- complete_data %>%
  group_by(pseudo_ID) %>%
  mutate(success_lag = lag(success, order_by = Date),
         Total.ST.min.lag = lag(Total.ST.min, order_by = Date),
         Proportion.ST_lag = lag(Proportion.ST, order_by = Date))


library(lme4)

model_4 <-
  glmer(
    Pickups ~   Pickups_lag_log +  Treatment + Phase + Proportion.ST +sex + Semester:Xt + devices + sex
    data = complete_data,
    family = poisson(link = "log"),
    control = glmerControl(optCtrl = list(maxfun = 1e5),optimizer = "bobyqa")
  )

summary(model_4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Pickups ~ Pickups_lag_log + Treatment + Phase + Proportion.ST +
##     sex + Semester:Xt + devices + sex:log(Total.ST.min.lag) +
##     (1 + Semester:Xt | pseudo_ID)
##    Data: complete_data
## Control: glmerControl(optCtrl = list(maxfun = 1e+05), optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  37635.5  37704.3 -18805.8  37611.5     2262
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.1710  -1.9936  -0.2117   1.6393  23.2599
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  pseudo_ID (Intercept) 0.1826   0.4273
##            Semester:Xt 0.0204   0.1428   -0.75
## Number of obs: 2274, groups:  pseudo_ID, 25
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 2.569914   0.227345  11.304  < 2e-16 ***
## Pickups_lag_log             0.319810   0.005826  54.892  < 2e-16 ***
## TreatmentB                 -0.111277   0.144824  -0.768   0.4423
## Phase                      -0.299695   0.010848 -27.628  < 2e-16 ***
## Proportion.ST               0.179393   0.014969  11.984  < 2e-16 ***
```

```
## sex                       -0.025511   0.131526  -0.194   0.8462
## devices                    0.135646   0.063878   2.124   0.0337 *
## Semester:Xt                0.149527   0.029061   5.145 2.67e-07 ***
## sex:log(Total.ST.min.lag)  0.003144   0.007547   0.417   0.6769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Pckp__ TrtmnB Phase  Prp.ST sex    devics Smst:X
## Pckps_lg_lg -0.099
## TreatmentB  -0.231  0.012
## Phase       -0.012  0.134  0.007
## Proportn.ST -0.029  0.021  0.005 -0.052
## sex         -0.401  0.025  0.160 -0.026 -0.023
## devices     -0.734 -0.022 -0.255 -0.006 -0.007  0.008
## Semester:Xt -0.278 -0.022 -0.019 -0.011 -0.009  0.006  0.008
## sx:(T.ST..)  0.009 -0.083  0.018  0.079  0.060 -0.326 -0.009 -0.003
```

AIC,BIC,Likelihood comparison of the model 1-4

```
AIC(model_1,model_2,model_3,model_4) %>% t() %>% as.data.frame() -> AIC_table
BIC(model_1,model_2,model_3,model_4) %>% t() %>% as.data.frame() -> BIC_table

rbind(AIC_table,BIC_table) %>% print()
```

```
##      model_1  model_2  model_3  model_4
## df     12.00    12.00    12.00    12.00
## AIC 37618.13 37628.01 37632.24 37635.52
## df1    12.00    12.00    12.00    12.00
## BIC 37686.88 37696.76 37700.99 37704.27
```

model 1 has the minimum AIC and BIC values. So, we will use model 1 for further analysis.

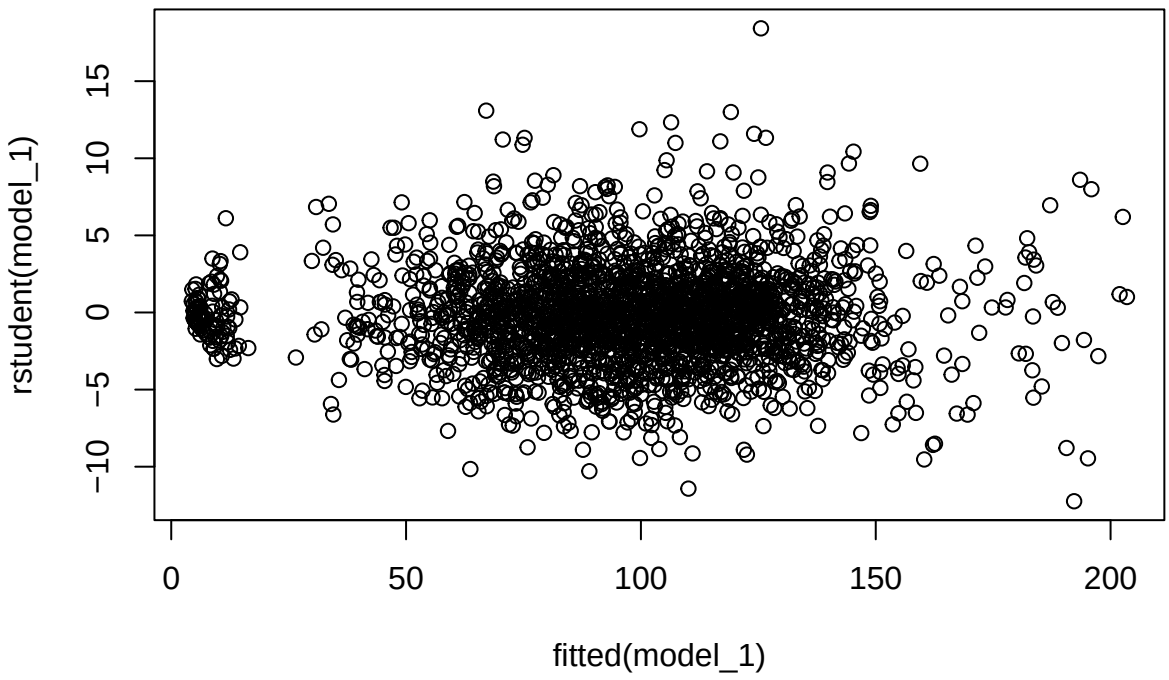
Checking Diagnostics

```
# plot(resid(model_1))

# plot the standardized residual

plot(fitted(model_1), rstudent(model_1 ))
```

```
## Warning in hatvalues.merMod(model): the hat matrix may not make sense for GLMMs
```




```
# without 7575

load("complete_data3_without_7575.RData")

complete_data  = complete_data %>% mutate(
  Pickups_lag_log = log(Pickups_lag),
  Total.ST.min.log = log(Total.ST.min)
)

# run under model 1


complete_data <- complete_data %>%
  group_by(pseudo_ID) %>%
  mutate(success_lag = lag(success, order_by = Date),
         Total.ST.min.lag = lag(Total.ST.min, order_by = Date),
         Proportion.ST_lag = lag(Proportion.ST, order_by = Date))


library(lme4)

model_1_without_7575 <-
  glmer(
    Pickups ~   Pickups_lag_log +  Treatment + Phase + Proportion.ST +sex + Semester:Xt + devices + sex
    data = complete_data,
    family = poisson(link = "log"),
```

```
    control = glmerControl(optCtrl = list(maxfun = 1e5),optimizer = "bobyqa")
  )

summary(model_1_without_7575)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Pickups ~ Pickups_lag_log + Treatment + Phase + Proportion.ST +
##     sex + Semester:Xt + devices + sex:log(Total.ST.min.lag) +
##     (1 + Semester:Xt | pseudo_ID)
##    Data: complete_data
## Control: glmerControl(optCtrl = list(maxfun = 1e+05), optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  32893.4  32960.5 -16434.7  32869.4     1969
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.0419  -2.0171  -0.1737   1.6874  15.7440
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  pseudo_ID (Intercept) 0.02796  0.1672
##            Semester:Xt 0.01795  0.1340   -0.44
## Number of obs: 1981, groups:  pseudo_ID, 22
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 3.159974   0.131256  24.075  < 2e-16 ***
## Pickups_lag_log             0.288159   0.006275  45.925  < 2e-16 ***
## TreatmentB                 -0.004416   0.085507  -0.052   0.9588
## Phase                      -0.314661   0.011551 -27.241  < 2e-16 ***
## Proportion.ST               0.152848   0.015376   9.941  < 2e-16 ***
## sex                        -0.072621   0.082608  -0.879   0.3793
## devices                    -0.016875   0.038733  -0.436   0.6631
## Semester:Xt                 0.142712   0.028977   4.925 8.44e-07 ***
## sex:log(Total.ST.min.lag)  0.016135   0.007839   2.058   0.0396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Pckp__ TrtmnB Phase  Prp.ST sex    devics Smst:X
## Pckps_lg_lg -0.217
## TreatmentB  -0.217  0.000
## Phase       -0.030  0.136  0.005
## Proportn.ST -0.068  0.020  0.003 -0.044
## sex         -0.250  0.042  0.012 -0.040 -0.018
## devices     -0.769  0.000 -0.240 -0.001 -0.004 -0.050
## Semester:Xt -0.119 -0.020 -0.006 -0.011 -0.006  0.002  0.005
## sx:(T.ST..) -0.015 -0.083  0.049  0.073  0.053 -0.545  0.010 -0.002
```

```
plot(fitted(model_1_without_7575), rstudent(model_1_without_7575 ))
```

```
## Warning in hatvalues.merMod(model): the hat matrix may not make sense for GLMMs
```