```r
library(readxl)
library(dplyr)
collegehookup <- read_excel("collegehookup.xlsx")
poisson.model <- glm(hookup_sum ~Gender+Age+Hisp+Black+Asian+BMI+
                       BMI2+college_dad + college_mom + hookup_highschool+
                       Siblings + ParentsDivorce,
                     collegehookup, family = poisson(link = "log"))
summary(poisson.model)
```

```
##
## Call:
## glm(formula = hookup_sum ~ Gender + Age + Hisp + Black + Asian +
##     BMI + BMI2 + college_dad + college_mom + hookup_highschool +
##     Siblings + ParentsDivorce, family = poisson(link = "log"),
##     data = collegehookup)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2481  -2.1380  -0.9925   0.7225   8.0514
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -6.956834   1.627297  -4.275 1.91e-05 ***
## Gender            -0.083049   0.075721  -1.097 0.272738
## Age               -0.019483   0.029417  -0.662 0.507771
## Hisp              -0.961821   0.156017  -6.165 7.06e-10 ***
## Black             -1.270728   0.381686  -3.329 0.000871 ***
## Asian              0.057045   0.095430   0.598 0.549990
## BMI                0.556374   0.133585   4.165 3.11e-05 ***
## BMI2              -0.010480   0.002853  -3.674 0.000239 ***
## college_dad        0.508710   0.162231   3.136 0.001714 **
## college_mom       -0.615654   0.119643  -5.146 2.66e-07 ***
## hookup_highschool  0.053597   0.003260  16.441  < 2e-16 ***
## Siblings           0.200127   0.032007   6.253 4.04e-10 ***
## ParentsDivorce     0.657869   0.101765   6.465 1.02e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1561.5  on 232  degrees of freedom
## Residual deviance: 1198.8  on 220  degrees of freedom
## AIC: 1754
##
## Number of Fisher Scoring iterations: 6
```

I think some of the coefficients can be explained as following:
1. The positive hookup_highschool coefficient is because if there is hookup in high school, students would be far more familiar with the process of hookup, making hookup much easier.
2. The negative college_mom coefficient is because if mother has a college degree, the control of parents toward the youngling would be stronger, which makes it harder for hookup.

3. The positive Siblings coefficient is because having more sibling can improve social ability, which makes it easier for hookup.
4. The positive ParentsDivorce coefficient is because the control of parent toward children would be alleviated if there is divorce.
The following is going to create the variable "peer_effect".

```
library(dplyr)
df <- data.frame(collegehookup)
df["peer_effect"] <- -1
df <- df[order(df$greek_group),]
this_row = 1
for( i in c(1:26)){
  temp_df <- filter(df, greek_group == i)
  if((!is.null(temp_df))){
    club_total_times = sum(temp_df$hookup_sum)
    club_menber_num = nrow(temp_df)
    if(club_menber_num>1){
      for( j in c(1:nrow(temp_df))){
        df[this_row,]["peer_effect"] <-
          (club_total_times-df[this_row,]["hookup_sum"])/(club_menber_num-1)
        this_row <- this_row+1
      }
    }
    else{
      this_row <- this_row+ club_menber_num
    }
  }
}
df<-df[!(df$peer_effect==-1 ),]
poisson.model2 <- glm(hookup_sum ~ peer_effect+Gender+Age+Hisp+Black+
                        Asian+BMI+BMI2+college_dad + college_mom +
                        hookup_highschool+Siblings + ParentsDivorce,
                      df, family = poisson(link = "log"))
summary(poisson.model2)
```

```
##
## Call:
## glm(formula = hookup_sum ~ peer_effect + Gender + Age + Hisp +
##     Black + Asian + BMI + BMI2 + college_dad + college_mom +
##     hookup_highschool + Siblings + ParentsDivorce, family = poisson(link = "log"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3596  -2.1458  -0.9472   0.7191   8.0360
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -7.400909   1.638097  -4.518 6.24e-06 ***
## peer_effect        0.014735   0.022250   0.662 0.507813
## Gender            -0.105705   0.078892  -1.340 0.180287
## Age               -0.006755   0.030168  -0.224 0.822832
## Hisp              -1.015350   0.161552  -6.285 3.28e-10 ***
## Black             -3.066706   1.012307  -3.029 0.002450 **
```

```
## Asian              0.057605   0.096134   0.599 0.549033
## BMI                0.568682   0.134301   4.234 2.29e-05 ***
## BMI2              -0.010699   0.002862  -3.739 0.000185 ***
## college_dad        0.481805   0.163773   2.942 0.003262 **
## college_mom       -0.605907   0.120757  -5.018 5.23e-07 ***
## hookup_highschool  0.054740   0.003288  16.649  < 2e-16 ***
## Siblings           0.203322   0.032545   6.247 4.17e-10 ***
## ParentsDivorce     0.646882   0.102187   6.330 2.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1536.3  on 225  degrees of freedom
## Residual deviance: 1158.4  on 212  degrees of freedom
## AIC: 1698.7
##
## Number of Fisher Scoring iterations: 6
```

When there is only one student in a club, there is no way to define peer_effect.
We remove the data from the dataset.
The p-value of peer_effect 0.507813.
As a result, we conclude that there is no peer effect.

```
summary(df$peer_effect)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.692   4.067   4.248   4.493  12.000
```