

MATH 523 Project

Jeremy Kim - 260620525

2017-04-28

Objective 1

```
table(fem,mar,anypub)
```

```
## , , anypub = No
##
##      mar
## fem  Married Single
##  Men      105     31
##  Women     71     68
##
```

```
## , , anypub = Yes
##
##      mar
## fem  Married Single
##  Men      276     82
##  Women    154    128
```

```
as.data.frame(table(fem,mar,anypub))
```

```
##      fem      mar anypub Freq
## 1  Men Married      No  105
## 2 Women Married      No   71
## 3  Men  Single      No   31
## 4 Women Single      No   68
## 5  Men Married     Yes  276
## 6 Women Married     Yes  154
## 7  Men  Single     Yes   82
## 8 Women Single     Yes  128
```

```
fem.count<-as.data.frame(table(fem,mar,anypub))$fem
mar.count<-as.data.frame(table(fem,mar,anypub))$mar
anypub.count<-as.data.frame(table(fem,mar,anypub))$anypub
count<-as.data.frame(table(fem,mar,anypub))$Freq
```

Fit saturated model first.

```
fit.fma<-glm(count~fem.count*mar.count*anypub.count,data=data,family="poisson")
summary(fit.fma)
```

```
##
## Call:
## glm(formula = count ~ fem.count * mar.count * anypub.count, family = "poisson",
##      data = data)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0  0
##
## Coefficients:
```

```
##                                Estimate Std. Error
## (Intercept)                   4.653960    0.097590
## fem.countWomen                -0.391280    0.153650
## mar.countSingle               -1.219973    0.204406
## anypub.countYes               0.966441    0.114660
## fem.countWomen:mar.countSingle 1.176801    0.265654
## fem.countWomen:anypub.countYes -0.192168    0.183644
## mar.countSingle:anypub.countYes 0.006292    0.240000
## fem.countWomen:mar.countSingle:anypub.countYes -0.148042    0.317327
##                                z value Pr(>|z|)
## (Intercept)                   47.689 < 2e-16 ***
## fem.countWomen                -2.547  0.0109 *
## mar.countSingle               -5.968 2.40e-09 ***
## anypub.countYes               8.429 < 2e-16 ***
## fem.countWomen:mar.countSingle 4.430 9.43e-06 ***
## fem.countWomen:anypub.countYes -1.046  0.2954
## mar.countSingle:anypub.countYes 0.026  0.9791
## fem.countWomen:mar.countSingle:anypub.countYes -0.467  0.6408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##    Null deviance: 3.1480e+02  on 7  degrees of freedom
## Residual deviance: 3.9968e-15  on 0  degrees of freedom
## AIC: 67.206
##
## Number of Fisher Scoring iterations: 3
```

High p value for 3-way interaction.

```
fit.fm.ma.fa<-update(fit.fma, ~.-fem.count:mar.count:anypub.count)
summary(fit.fm.ma.fa)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count +
##      fem.count:mar.count + fem.count:anypub.count + mar.count:anypub.count,
##      family = "poisson", data = data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## 0.14498 -0.17427 -0.26145  0.18062 -0.08884  0.11953  0.16427
##      8
## -0.13019
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   4.63978    0.09340 49.679 < 2e-16 ***
## fem.countWomen                -0.35649    0.13391 -2.662  0.00776 **
## mar.countSingle               -1.15920    0.15498 -7.480 7.45e-14 ***
## anypub.countYes               0.98597    0.10722  9.195 < 2e-16 ***
## fem.countWomen:mar.countSingle 1.07343    0.14520  7.393 1.44e-13 ***
## fem.countWomen:anypub.countYes -0.24195    0.14923 -1.621  0.10494
## mar.countSingle:anypub.countYes -0.07802    0.15636 -0.499  0.61778
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 314.80384  on 7  degrees of freedom
## Residual deviance:   0.21849  on 1  degrees of freedom
## AIC: 65.425
##
## Number of Fisher Scoring iterations: 3
```

Both mar:anypub and fem:anypub have high p value.

```
fit.fm.ma<-update(fit.fm.ma.fa, ~.- fem.count:anypub.count)
summary(fit.fm.ma)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count +
##      fem.count:mar.count + mar.count:anypub.count, family = "poisson",
##      data = data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
## -0.5421  0.6896 -0.8869  0.6479  0.3427 -0.4501  0.5873 -0.4539
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.70640    0.08159  57.686 < 2e-16 ***
## fem.countWomen    -0.52670    0.08408  -6.264 3.74e-10 ***
## mar.countSingle   -1.11724    0.14957  -7.470 8.03e-14 ***
## anypub.countYes    0.89330    0.08948   9.983 < 2e-16 ***
## fem.countWomen:mar.countSingle  1.07743    0.14499   7.431 1.08e-13 ***
## mar.countSingle:anypub.countYes -0.14131    0.15123  -0.934  0.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 314.8038  on 7  degrees of freedom
## Residual deviance:   2.8469  on 2  degrees of freedom
## AIC: 66.053
##
## Number of Fisher Scoring iterations: 4
```

```
fit.fm.fa<-update(fit.fm.ma.fa, ~.- mar.count:anypub.count)
summary(fit.fm.ma)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count +
##      fem.count:mar.count + mar.count:anypub.count, family = "poisson",
##      data = data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
```

```
## -0.5421  0.6896 -0.8869  0.6479  0.3427 -0.4501  0.5873 -0.4539
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.70640    0.08159  57.686 < 2e-16 ***
## fem.countWomen   -0.52670    0.08408  -6.264 3.74e-10 ***
## mar.countSingle  -1.11724    0.14957  -7.470 8.03e-14 ***
## anypub.countYes   0.89330    0.08948   9.983 < 2e-16 ***
## fem.countWomen:mar.countSingle  1.07743    0.14499   7.431 1.08e-13 ***
## mar.countSingle:anypub.countYes -0.14131    0.15123  -0.934  0.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 314.8038 on 7 degrees of freedom
## Residual deviance:  2.8469 on 2 degrees of freedom
## AIC: 66.053
##
## Number of Fisher Scoring iterations: 4
```

mar:anypub has high p value.

```
fit.fm.a<-update(fit.fm.fa, ~. - fem.count:anypub.count)
summary(fit.fm.a)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count +
##      fem.count:mar.count, family = "poisson", data = data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
## -0.9013  0.4073 -0.5159  1.1561  0.5790 -0.2702  0.3311 -0.7854
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.74065    0.07189  65.943 < 2e-16 ***
## fem.countWomen   -0.52670    0.08408  -6.264 3.74e-10 ***
## mar.countSingle  -1.21541    0.10712 -11.346 < 2e-16 ***
## anypub.countYes   0.84470    0.07210  11.715 < 2e-16 ***
## fem.countWomen:mar.countSingle  1.07743    0.14499   7.431 1.08e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 314.8038 on 7 degrees of freedom
## Residual deviance:  3.7156 on 3 degrees of freedom
## AIC: 64.922
##
## Number of Fisher Scoring iterations: 4
```

```
fit.f.m.a<-update(fit.fm.a, ~. - fem.count:mar.count)
summary(fit.f.m.a)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count, family = "poisson",
##      data = data)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
## 0.6651 -1.4364 -2.9099  3.5562  3.0185 -3.0510 -3.3939  2.7407
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.58834    0.07159  64.093   <2e-16 ***
## fem.countWomen -0.15990    0.06633  -2.411    0.0159 *
## mar.countSingle -0.67354    0.06990  -9.635   <2e-16 ***
## anypub.countYes  0.84470    0.07210  11.715   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 314.80  on 7  degrees of freedom
## Residual deviance:  61.07  on 4  degrees of freedom
## AIC: 120.28
##
## Number of Fisher Scoring iterations: 4
```

Compare models against saturated model using goodness of fit tests.

```
modellist = list(fit.f.m.a, fit.fm.a, fit.fm.fa, fit.fm.ma, fit.fm.ma.fa, fit.fma)

data.frame(Model=c("f.m.a", "fm.a", "fm.fa", "fm.ma", "fm.ma.fa", "fma"), Dev=round(unlist(lapply(modellist,
  X2=round(unlist(lapply(modellist, function(x){ sum( residuals(x,type="pearson")^2)})),4),
  AIC=round(unlist(lapply(modellist, AIC)),4),
  BIC=round(unlist(lapply(modellist, AIC,k=log(915)))),4),
  Df = unlist(lapply(modellist,function(x){ x$df.residual})))
```

| ## | Model | Dev | X2 | AIC | BIC | Df |
|------|----------|---------|---------|----------|----------|----|
| ## 1 | f.m.a | 61.0702 | 61.5172 | 120.2762 | 139.5519 | 4 |
| ## 2 | fm.a | 3.7156 | 3.7458 | 64.9217 | 89.0163 | 3 |
| ## 3 | fm.fa | 0.4668 | 0.4672 | 63.6728 | 92.5864 | 2 |
| ## 4 | fm.ma | 2.8469 | 2.8315 | 66.0529 | 94.9665 | 2 |
| ## 5 | fm.ma.fa | 0.2185 | 0.2177 | 65.4245 | 99.1570 | 1 |
| ## 6 | fma | 0.0000 | 0.0000 | 67.2060 | 105.7574 | 0 |

There is evidence against f.m.a and fm.ma. Model fm.a seems to perform the best.

```
anova(fit.fm.a, fit.fm.fa, fit.fm.ma.fa, fit.fma,test="LRT")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: count ~ fem.count + mar.count + anypub.count + fem.count:mar.count
```

```
## Model 2: count ~ fem.count + mar.count + anypub.count + fem.count:mar.count +
##      fem.count:anypub.count
```

```
## Model 3: count ~ fem.count + mar.count + anypub.count + fem.count:mar.count +
##      fem.count:anypub.count + mar.count:anypub.count
```

```
## Model 4: count ~ fem.count * mar.count * anypub.count
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         3      3.7156
## 2         2      0.4668 1   3.2488 0.07147 .
## 3         1      0.2185 1   0.2483 0.61826
## 4         0      0.0000 1   0.2185 0.64020
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model fm.a is the most suitable model. Thus, there are more biochemists who have published an article, are men, and are married.

```
summary(fit.fm.a)
```

```
##
## Call:
## glm(formula = count ~ fem.count + mar.count + anypub.count +
##      fem.count:mar.count, family = "poisson", data = data)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -0.9013  0.4073 -0.5159  1.1561  0.5790 -0.2702  0.3311 -0.7854
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.74065    0.07189  65.943 < 2e-16 ***
## fem.countWomen    -0.52670    0.08408  -6.264 3.74e-10 ***
## mar.countSingle   -1.21541    0.10712 -11.346 < 2e-16 ***
## anypub.countYes    0.84470    0.07210  11.715 < 2e-16 ***
## fem.countWomen:mar.countSingle 1.07743    0.14499   7.431 1.08e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##   Null deviance: 314.8038  on 7  degrees of freedom
## Residual deviance:   3.7156  on 3  degrees of freedom
## AIC: 64.922
##
## Number of Fisher Scoring iterations: 4
```

```
confint(fit.fm.a)
```

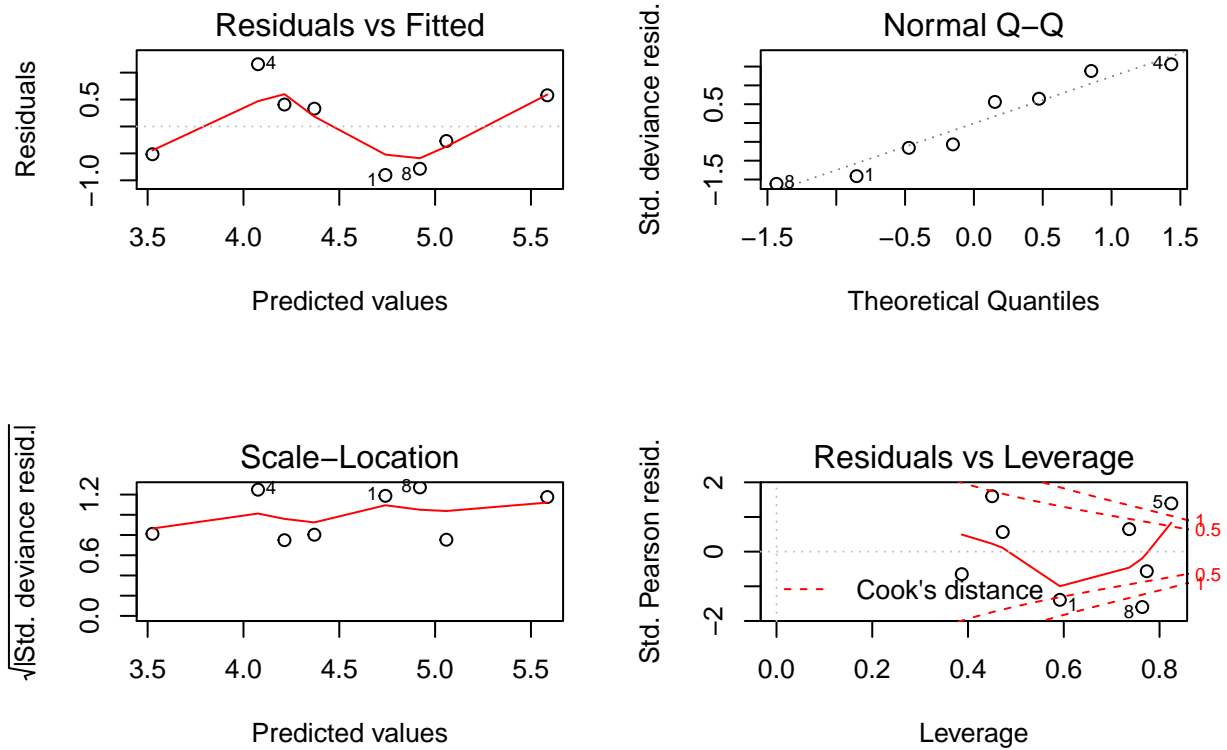
```
## Waiting for profiling to be done...
##
##              2.5 %      97.5 %
## (Intercept)    4.5973660  4.8792318
## fem.countWomen -0.6927523 -0.3629811
## mar.countSingle -1.4295664 -1.0092477
## anypub.countYes  0.7046459  0.9874093
## fem.countWomen:mar.countSingle 0.7950849  1.3637464
```

```
exp(confint(fit.fm.a))
```

```
## Waiting for profiling to be done...
##
##              2.5 %      97.5 %
## (Intercept)  99.2226224 131.5295878
## fem.countWomen  0.5001975  0.6955996
```

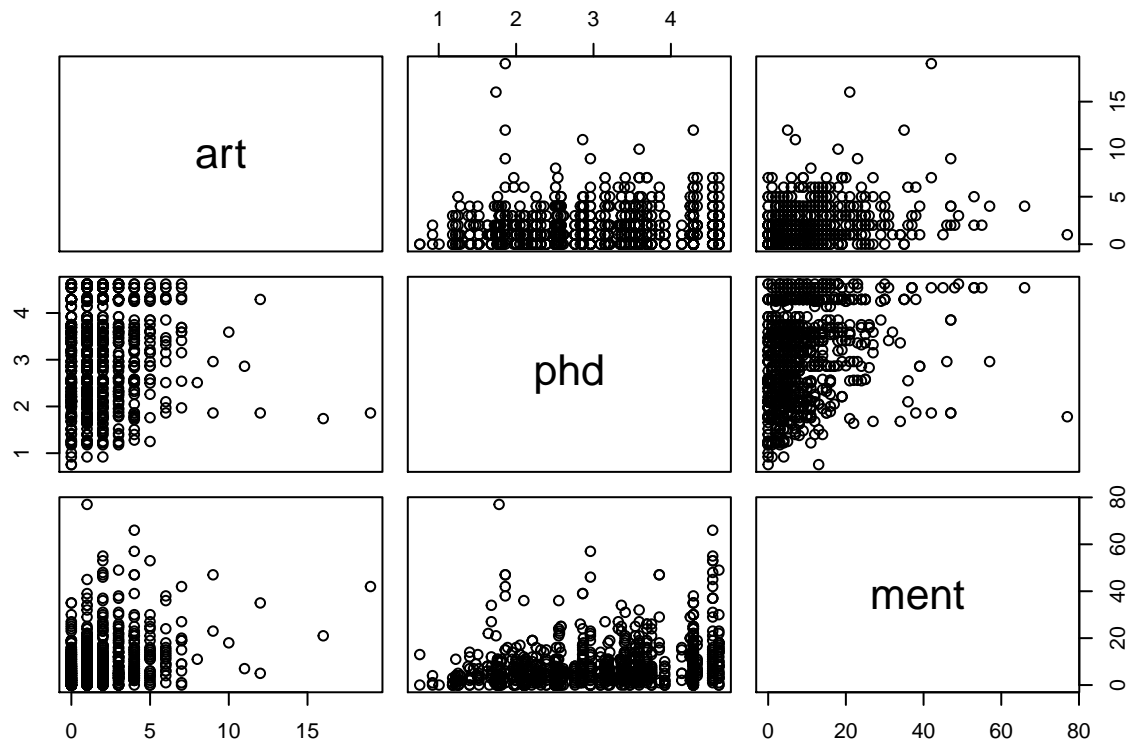
```
## mar.countSingle      0.2394127  0.3644931
## anypub.countYes      2.0231302  2.6842714
## fem.countWomen:mar.countSingle 2.2146290  3.9108175
```

```
par(mfrow=c(2,2))
plot(fit.fm.a)
```



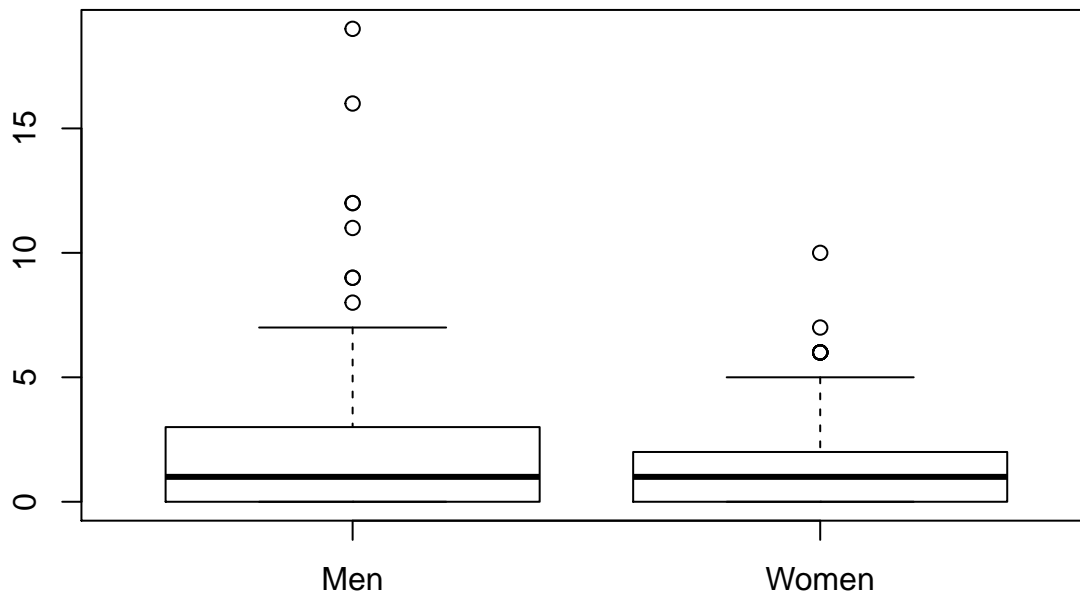
Objective 2

```
par(mfrow=c(1,1))
pairs(cbind(art, phd, ment))
```



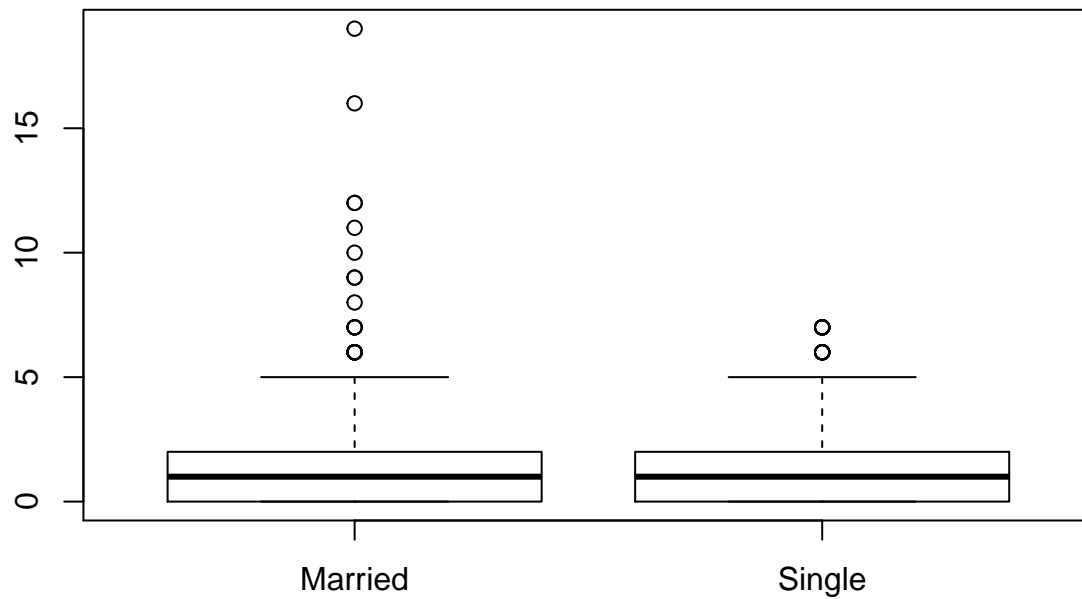
art and phd seems to have no significant correlation, though phd and ment seem to do.

```
boxplot(art~fem)
```



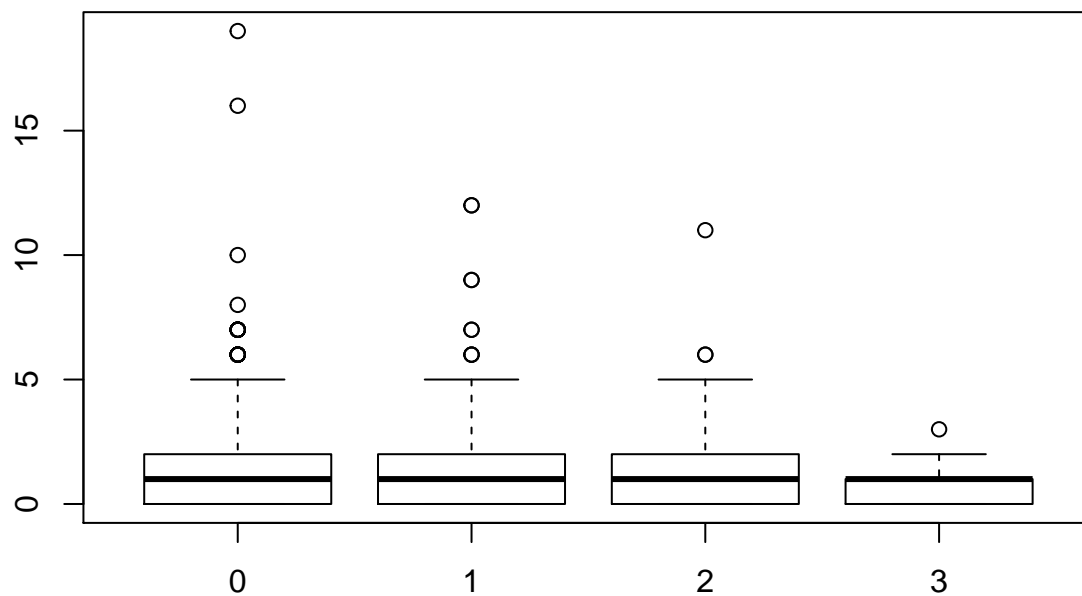
Men tend to publish more articles.

```
boxplot(art~mar)
```

Married people tend to publish more articles. The assumption seems fair as married person tends to be older, thus more experienced and renown.

```
boxplot(art~kid5)
```



People with less kids tend to publish more articles. The assumption seems fair as more kids means less time to spend on research.

Let's fit all covariates without interaction.

```
fit.f.m.k.p.mt<-glm(art ~ fem + mar + kid5 + phd + ment, family = poisson, data = data)
summary(fit.f.m.k.p.mt)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment, family = poisson,
##      data = data)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5672  -1.5398  -0.3660   0.5722   5.4467
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.459860   0.093335   4.927 8.35e-07 ***
## femWomen    -0.224594   0.054613  -4.112 3.92e-05 ***
## marSingle   -0.155243   0.061374  -2.529  0.0114 *
## kid5        -0.184883   0.040127  -4.607 4.08e-06 ***
## phd          0.012823   0.026397   0.486  0.6271
## ment         0.025543   0.002006  12.733 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1634.4  on 909  degrees of freedom
## AIC: 3314.1
##
## Number of Fisher Scoring iterations: 5
```

Add interaction between phd and mentor, as they seems to have a significant correlation.

```
fit.f.m.k.pmt<-glm(art ~ fem + mar + kid5 + phd * ment, family = poisson, data = data)
summary(fit.f.m.k.pmt)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd * ment, family = poisson,
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4416  -1.5493  -0.3727   0.5594   5.3927
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.313675   0.115110   2.725  0.00643 **
## femWomen    -0.224786   0.054666  -4.112 3.92e-05 ***
## marSingle   -0.149975   0.061458  -2.440  0.01468 *
## kid5        -0.182234   0.040215  -4.532 5.86e-06 ***
## phd          0.061286   0.034355   1.784  0.07444 .
## ment         0.037267   0.005503   6.772 1.27e-11 ***
## phd:ment     -0.003722   0.001676  -2.221  0.02637 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1629.5  on 908  degrees of freedom
## AIC: 3311.2
##
```

```
## Number of Fisher Scoring iterations: 5
```

Take out phd from the first model, as it has high p value.

```
fit.f.m.k.mt<-update(fit.f.m.k.p.mt, ~. - phd)
summary(fit.f.m.k.mt)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + ment, family = poisson,
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6436  -1.5408  -0.3583   0.5623   5.3986
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.49735    0.05224   9.520 < 2e-16 ***
## femWomen     -0.22530    0.05461  -4.125 3.70e-05 ***
## marSingle    -0.15218    0.06107  -2.492  0.0127 *
## kid5         -0.18499    0.04014  -4.609 4.05e-06 ***
## ment         0.02576    0.00195  13.212 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1634.6  on 910  degrees of freedom
## AIC: 3312.3
##
## Number of Fisher Scoring iterations: 5
```

Further reduce the model by taking out mar.

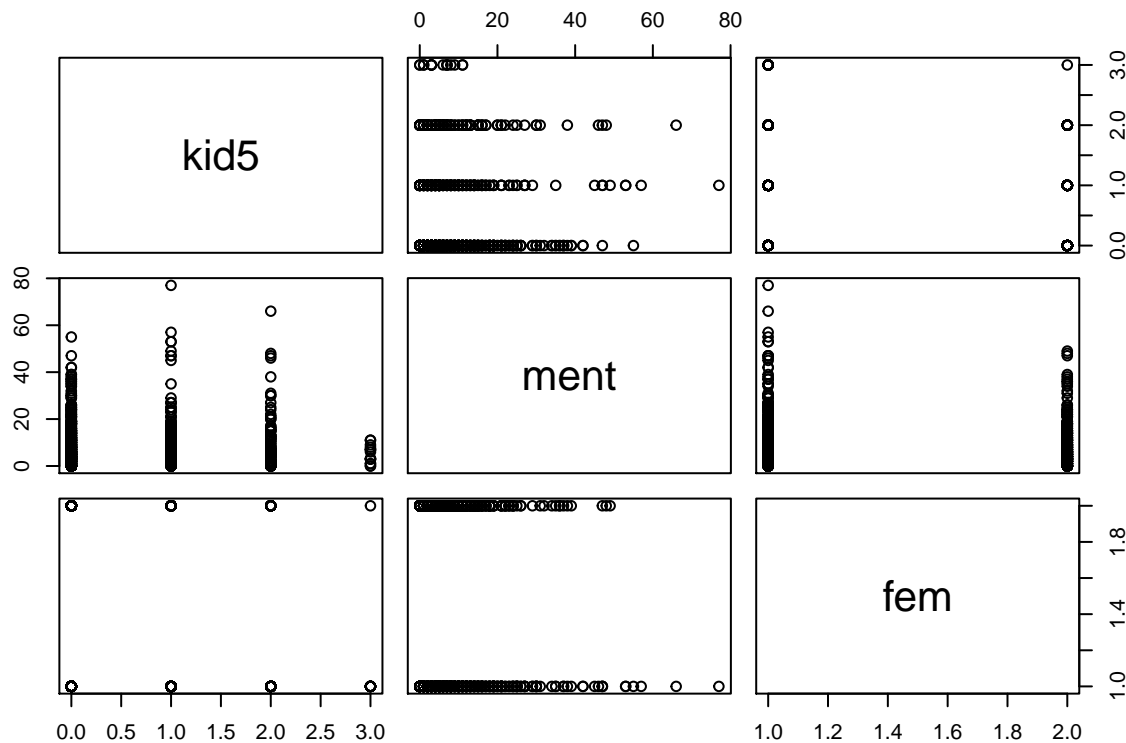
```
fit.f.k.mt<-update(fit.f.m.k.mt, ~. - mar)
summary(fit.f.k.mt)
```

```
##
## Call:
## glm(formula = art ~ fem + kid5 + ment, family = poisson, data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5746  -1.5633  -0.3652   0.5583   5.5514
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.436733    0.046825   9.327 < 2e-16 ***
## femWomen     -0.243106    0.054207  -4.485 7.30e-06 ***
## kid5         -0.143787    0.036573  -3.931 8.44e-05 ***
## ment         0.025645    0.001952  13.136 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 1817.4 on 914 degrees of freedom
## Residual deviance: 1640.9 on 911 degrees of freedom
## AIC: 3316.6
##
## Number of Fisher Scoring iterations: 5
```

Add more interaction to model f.m.k.pmt.

```
pairs(cbind(kid5, ment, fem))
```



kid5 and ment have some correlation. Let's add them.

```
fit.f.m.kmt.pmt<-update(fit.f.m.k.pmt, ~. + kid5:ment)
summary(fit.f.m.kmt.pmt)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment + phd:ment +
##      kid5:ment, family = poisson, data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2304  -1.5504  -0.3601   0.5661   5.3625
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.277368  0.117517   2.360  0.0183 *
## femWomen    -0.218480  0.054784  -3.988 6.66e-05 ***
## marSingle   -0.155186  0.061502  -2.523  0.0116 *
## kid5        -0.120575  0.051046  -2.362  0.0182 *
```

```
## phd          0.059428  0.034598  1.718  0.0859 .
## ment         0.041681  0.006106  6.826 8.71e-12 ***
## phd:ment     -0.003858  0.001717 -2.247  0.0246 *
## kid5:ment    -0.005451  0.002874 -1.897  0.0578 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 1817.4 on 914 degrees of freedom
## Residual deviance: 1625.8 on 907 degrees of freedom
## AIC: 3309.5
##
## Number of Fisher Scoring iterations: 5
```

Compare models using goodness of fit tests.

```
modellist2 = list(fit.f.k.mt, fit.f.m.k.mt, fit.f.m.k.p.mt, fit.f.m.k.pmt, fit.f.m.kmt.pmt)
data.frame(Model=c("f.k.mt", "f.m.k.mt", "f.m.k.p.mt", "f.m.k.pmt", "f.m.kmt.pmt"), Dev=round(unlist(lapply(
  X2=round(unlist(lapply(modellist2, function(x){ sum( residuals(x,type="pearson")^2)})),4),
  AIC=round(unlist(lapply(modellist2, AIC)),4),
  BIC=round(unlist(lapply(modellist2, AIC,k=log(915))),4),
  Df = unlist(lapply(modellist2,function(x){ x$df.residual})))
```

```
##      Model      Dev      X2      AIC      BIC Df
## 1      f.k.mt 1640.851 1668.842 3316.593 3335.869 911
## 2      f.m.k.mt 1634.607 1659.987 3312.349 3336.443 910
## 3      f.m.k.p.mt 1634.371 1662.547 3314.113 3343.026 909
## 4      f.m.k.pmt 1629.493 1645.090 3311.235 3344.967 908
## 5      f.m.kmt.pmt 1625.805 1633.618 3309.547 3348.099 907
```

```
anova(fit.f.k.mt, fit.f.m.k.mt, fit.f.m.k.p.mt, fit.f.m.k.pmt, fit.f.m.kmt.pmt,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: art ~ fem + kid5 + ment
## Model 2: art ~ fem + mar + kid5 + ment
## Model 3: art ~ fem + mar + kid5 + phd + ment
## Model 4: art ~ fem + mar + kid5 + phd * ment
## Model 5: art ~ fem + mar + kid5 + phd + ment + phd:ment + kid5:ment
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      911      1640.8
## 2      910      1634.6  1    6.2442  0.01246 *
## 3      909      1634.4  1    0.2362  0.62696
## 4      908      1629.5  1    4.8779  0.02720 *
## 5      907      1625.8  1    3.6876  0.05482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model f.m.kmt.pmt seems to be the best fit.

```
X2 =sum(residuals(fit.f.m.kmt.pmt,type="pearson")^2)
cat(c("X2 = ",round(X2,3)))
```

```
## X2 = 1633.618
```

```
cat(c("Rejection value = ",round(qchisq(0.95,914),3)))
```

```

## Rejection value = 985.444
phi = X2/(907)
cat(c("Phi=",round(phi,3)))

## Phi= 1.801
Phi is 1.8, let's see how negative binomial model performs.
fit.f.m.kmt.pmt.nb<-glm.nb(art ~ fem + mar + kid5 + phd + ment + phd:ment + kid5:ment, data = data)
summary(fit.f.m.kmt.pmt.nb)

##
## Call:
## glm.nb(formula = art ~ fem + mar + kid5 + phd + ment + phd:ment +
##      kid5:ment, data = data, init.theta = 2.305959843, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6587  -1.3723  -0.2781   0.4434   3.5130
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.183703   0.157604   1.166  0.24378
## femWomen     -0.215691   0.072523  -2.974  0.00294 **
## marSingle    -0.147076   0.081954  -1.795  0.07272 .
## kid5         -0.136905   0.066916  -2.046  0.04076 *
## phd           0.078903   0.046572   1.694  0.09022 .
## ment          0.053060   0.010350   5.126 2.95e-07 ***
## phd:ment     -0.006459   0.002874  -2.247  0.02461 *
## kid5:ment    -0.003656   0.004178  -0.875  0.38155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.306) family taken to be 1)
##
##      Null deviance: 1115.7  on 914  degrees of freedom
## Residual deviance: 1005.0  on 907  degrees of freedom
## AIC: 3134.7
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  2.306
##              Std. Err.:  0.279
##
## 2 x log-likelihood: -3116.702
fit.f.m.kmt.pmt.quasi<-glm(art ~ fem + mar + kid5 + phd + ment + phd:ment + kid5:ment,family=quasi(link=
summary(fit.f.m.kmt.pmt.quasi)

##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment + phd:ment +
##      kid5:ment, family = quasi(link = "log", variance = "mu"),
##      data = data)

```

```

##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2304  -1.5504  -0.3601   0.5661   5.3625
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.277368   0.157716   1.759  0.07897 .
## femWomen    -0.218480   0.073523  -2.972  0.00304 **
## marSingle   -0.155186   0.082540  -1.880  0.06041 .
## kid5        -0.120575   0.068507  -1.760  0.07874 .
## phd          0.059428   0.046433   1.280  0.20091
## ment        0.041681   0.008195   5.086 4.43e-07 ***
## phd:ment    -0.003858   0.002305  -1.674  0.09442 .
## kid5:ment   -0.005451   0.003857  -1.413  0.15786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasi family taken to be 1.801149)
##
##      Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1625.8  on 907  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
fit.f.m.kmt.pmt.quasi2<-glm(art ~ fem + mar + kid5 + phd + ment + phd:ment + kid5:ment,family=quasi(link=
summary(fit.f.m.kmt.pmt.quasi2)

##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment + phd:ment +
##      kid5:ment, family = quasi(link = "log", variance = "mu^2"),
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2318  -0.1904   0.0000   0.4438   3.2294
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.141323   0.161188   0.877  0.38085
## femWomen    -0.211595   0.074983  -2.822  0.00488 **
## marSingle   -0.152154   0.085107  -1.788  0.07414 .
## kid5        -0.144861   0.066833  -2.168  0.03046 *
## phd          0.083843   0.047922   1.750  0.08053 .
## ment        0.060275   0.012942   4.657 3.68e-06 ***
## phd:ment    -0.007730   0.003562  -2.170  0.03027 *
## kid5:ment   -0.003486   0.004825  -0.723  0.47016
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasi family taken to be 1.134446)
##
##      Null deviance: 368.50  on 914  degrees of freedom

```

```
## Residual deviance: 361.85 on 907 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 16
```

Neither negative binomial nor quasi-poisson performs well. Thus, the poisson model f.m.kmt.pmt is the most suitable model.

```
summary(fit.f.m.kmt.pmt)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment + phd:ment +
##      kid5:ment, family = poisson, data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2304 -1.5504 -0.3601  0.5661  5.3625
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.277368   0.117517   2.360   0.0183 *
## femWomen     -0.218480   0.054784  -3.988 6.66e-05 ***
## marSingle    -0.155186   0.061502  -2.523   0.0116 *
## kid5         -0.120575   0.051046  -2.362   0.0182 *
## phd           0.059428   0.034598   1.718   0.0859 .
## ment          0.041681   0.006106   6.826 8.71e-12 ***
## phd:ment     -0.003858   0.001717  -2.247   0.0246 *
## kid5:ment    -0.005451   0.002874  -1.897   0.0578 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1817.4 on 914 degrees of freedom
## Residual deviance: 1625.8 on 907 degrees of freedom
## AIC: 3309.5
##
## Number of Fisher Scoring iterations: 5
```

```
confint(fit.f.m.kmt.pmt)
```

```
## Waiting for profiling to be done...
##
##              2.5 %      97.5 %
## (Intercept)  0.045126131  0.5058226637
## femWomen     -0.326150959 -0.1113626905
## marSingle    -0.276036961 -0.0348891449
## kid5         -0.221562327 -0.0214366335
## phd          -0.008298593  0.1273326456
## ment         0.029571614  0.0535420729
## phd:ment     -0.007219907 -0.0004809381
## kid5:ment    -0.011158832  0.0001111235
```

```
exp(confint(fit.f.m.kmt.pmt))
```

```
## Waiting for profiling to be done...
```



```
##                2.5 %    97.5 %
## (Intercept) 1.0461598 1.6583492
## femWomen    0.7216962 0.8946142
## marSingle   0.7587849 0.9657125
## kid5        0.8012660 0.9787915
## phd         0.9917357 1.1357948
## ment        1.0300132 1.0550014
## phd:ment    0.9928061 0.9995192
## kid5:ment   0.9889032 1.0001111
```

The prestige of Ph.D. program and the number of publications by Ph.D. mentor seem to positively affect the number of articles published, though with a slightly negative interaction (less “synergy” effect as the phd and ment increase). Being single seem to negatively affect the number of articles, as singles tend to be young. Having more kids and being female also seem to negatively affect the number of articles.

```
par(mfrow=c(2,2))
plot(fit.f.m.kmt.pmt)
```

