

hw10

Mohammad Mottahedi

April 4, 2016

```
set.seed(1234)
library(vcd)
```

```
## Loading required package: grid
```

```
setwd('~/.Dropbox/spring2016/discretDataAnalysis/hw_solutions/hw10/')
```

Problem 1

part a

$$\log(\mu_{ij}) = \lambda + \lambda_i^{Inc} + \lambda_j^{Pol}$$

$$\log(\mu_{11}) = \lambda + \lambda_1^{Inc} + \lambda_1^{Pol}$$

```
montana <- read.csv("montana.csv", header = T, na.strings = ".")
montana <- na.omit(montana)
attach(montana)
```

```
inc_pol <- xtabs(data=montana, ~Inc+Pol)
inc_pol
```

```
##      Pol
## Inc  1  2  3
##   1 20  6 14
##   2 23 21 26
##   3 23  7 23
```

```
model <- loglin(inc_pol, list(c(1,2)), fit=T, param=T)
```

```
## 2 iterations: deviation 0
```

```
#model
inc_pol <- as.data.frame(inc_pol)
model <- glm(Freq ~ Inc + Pol, data = inc_pol, family = poisson())
summary(model)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol, family = poisson(), data = inc_pol)
##
## Deviance Residuals:
```

```
##      1      2      3      4      5      6      7      8
## 0.9114 -1.0380 0.3286 -0.8546 1.5703 -1.3086 -0.3774 -0.2042
##      9
## 0.5449
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.78478    0.18444  15.099 < 2e-16 ***
## Inc2         0.55962    0.19821   2.823 0.00475 **
## Inc3         0.28141    0.20945   1.344 0.17908
## Pol2        -0.66329    0.21110  -3.142 0.00168 **
## Pol3        -0.04652    0.17614  -0.264 0.79169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 28.1886  on 8  degrees of freedom
## Residual deviance:  7.4059  on 4  degrees of freedom
## AIC: 59.136
##
## Number of Fisher Scoring iterations: 4
```

```
# fit
1 - pchisq(model$deviance, 1)
```

```
## [1] 0.006500918
```

p-value is < 0.05 we can reject the null hypothesis of independence.

$$\log(\mu_{11}) = \lambda + \lambda_1^{Inc} + \lambda_1^{Pol} = 2.78 + 0 + 0 = 2.78$$

$$\mu_{11} = e^{2.78} = 16.1190209$$

Since we rejected the null hypothesis of independence we can say that income and political affiliation are related.

part b

We ignored other variables in the dataset.

part c

$$\log(\mu_{ij}) = \lambda + \lambda_i^{Inc} + \lambda_j^{Pol} + \lambda_{ij}^{Pol,Inc}$$

```
sat.model <- glm(Freq ~ Inc + Pol + Inc * Pol ,
                 data = inc_pol, family = poisson())
summary(sat.model)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol + Inc * Pol, family = poisson(),
```

```
##      data = inc_pol)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.99573    0.22361  13.397 < 2e-16 ***
## Inc2         0.13976    0.30574   0.457  0.64758
## Inc3         0.13976    0.30574   0.457  0.64758
## Pol2        -1.20397    0.46547  -2.587  0.00969 **
## Pol3        -0.35667    0.34847  -1.024  0.30604
## Inc2:Pol2     1.11300    0.55476   2.006  0.04483 *
## Inc3:Pol2     0.01439    0.63482   0.023  0.98192
## Inc2:Pol3     0.47928    0.45096   1.063  0.28788
## Inc3:Pol3     0.35667    0.45649   0.781  0.43460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2.8189e+01  on 8  degrees of freedom
## Residual deviance: 1.7764e-15  on 0  degrees of freedom
## AIC: 59.73
##
## Number of Fisher Scoring iterations: 3
```

The saturated model is the most complex model, the residual deviance is almost 0 and the model fits perfectly to the observed data, but all the interaction term in the saturated model are insignificant (p-value >0.5) except one. We can pick the independence model.

calculated odds ratio:

$$\log\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) = \lambda_{11}^{Inc,Pol} + \lambda_{22}^{Inc,Pol} - \lambda_{12}^{Inc,Pol} - \lambda_{21}^{Pol,Inc} = 1.113, \exp() = 3.0434751$$

$$\log\left(\frac{\mu_{12}\mu_{23}}{\mu_{13}\mu_{22}}\right) = \lambda_{12}^{Inc,Pol} + \lambda_{23}^{Inc,Pol} - \lambda_{13}^{Inc,Pol} - \lambda_{22}^{Pol,Inc} = -0.63372, \exp() = 0.5306142$$

$$\log\left(\frac{\mu_{21}\mu_{32}}{\mu_{31}\mu_{22}}\right) = \lambda_{21}^{Inc,Pol} + \lambda_{32}^{Inc,Pol} - \lambda_{31}^{Inc,Pol} - \lambda_{22}^{Pol,Inc} = -1.09861, \exp() = 0.3333341$$

$$\log\left(\frac{\mu_{22}\mu_{33}}{\mu_{23}\mu_{32}}\right) = \lambda_{22}^{Inc,Pol} + \lambda_{33}^{Inc,Pol} - \lambda_{23}^{Inc,Pol} - \lambda_{32}^{Pol,Inc} = 0.976, \exp() = 2.6538197$$

They are approximately equal to the odds ratios calculated based on the observed data. Since only

part d

odds ratio from hw5

```
inc_pol <- xtabs(data=montana, ~Inc+Pol)
oddsratio(inc_pol, log = F)
```

```
## odds ratios for Inc and Pol
##
##      Pol
## Inc      1:2      2:3
## 1:2 3.0434783 0.5306122
## 2:3 0.3333333 2.6538462
```

```
exp(confint(oddsratio(inc_pol)))
```

```
##           2.5 %    97.5 %
## 1:2/1:2 1.0260146 9.0279023
## 2:3/1:2 0.1187230 0.9358853
## 1:2/2:3 0.1738335 1.6196498
## 2:3/2:3 0.9541222 7.3815488
```

since the indendence model fits better, we reach the same conclusion as in hw5. the income and political affiliation seems to be unrelated.

part e

```
inc_pol_sex <- xtabs(data=montana, ~Inc+Pol+Sex)
#model
inc_pol <- as.data.frame(inc_pol)
model <- glm(Freq ~ Inc + Pol + Sex, data = inc_pol_sex, family = poisson())
summary(model)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol + Sex, family = poisson(), data = inc_pol_sex)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14358  -0.67414  -0.00218   0.29702   2.04041
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.20189    0.19716  11.168 < 2e-16 ***
## Inc2         0.55962    0.19821   2.823  0.00475 **
## Inc3         0.28141    0.20945   1.344  0.17908
## Pol2        -0.66329    0.21110  -3.142  0.00168 **
## Pol3        -0.04652    0.17614  -0.264  0.79169
## Sex1        -0.23419    0.15773  -1.485  0.13760
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 46.928  on 17  degrees of freedom
## Residual deviance: 23.925  on 12  degrees of freedom
## AIC: 105.7
##
## Number of Fisher Scoring iterations: 5
```

```
homo <- glm(Freq ~ Inc + Pol + Sex + Sex*Inc + Pol*Inc + Sex*Pol, data = inc_pol_sex, family = poisson)
summary(homo)
```

```
##
```

```
## Call:
## glm(formula = Freq ~ Inc + Pol + Sex + Sex * Inc + Pol * Inc +
##      Sex * Pol, family = poisson(), data = inc_pol_sex)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
## 0.1956 0.4172 -0.4858 -0.5974 0.1339 0.2062 0.1790 -0.4787
##      9     10     11     12     13     14     15     16
## 0.3150 -0.1383 -0.3696 0.6188 0.5296 -0.1654 -0.4279 -0.1713
##     17     18
## 0.5253 -0.6072
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.871043   0.336864   5.554 2.79e-08 ***
## Inc2          0.398363   0.403166   0.988 0.32311
## Inc3          0.825696   0.397353   2.078 0.03771 *
## Pol2         -0.783210   0.531944  -1.472 0.14093
## Pol3          0.006451   0.415846   0.016 0.98762
## Sex1          0.732012   0.373119   1.962 0.04978 *
## Inc2:Sex1     -0.411674   0.411451  -1.001 0.31705
## Inc3:Sex1     -1.328440   0.449442  -2.956 0.00312 **
## Inc2:Pol2      1.041380   0.560211   1.859 0.06304 .
## Inc3:Pol2     -0.207133   0.653622  -0.317 0.75132
## Inc2:Pol3      0.418991   0.456067   0.919 0.35825
## Inc3:Pol3      0.168092   0.474655   0.354 0.72324
## Pol2:Sex1     -0.710568   0.452444  -1.571 0.11630
## Pol3:Sex1     -0.599562   0.371601  -1.613 0.10665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 46.9276  on 17  degrees of freedom
## Residual deviance:  2.9298  on  4  degrees of freedom
## AIC: 100.71
##
## Number of Fisher Scoring iterations: 4
```

$$\Delta G^2 = 20.9902, \Delta df = 8$$

p-value: 0.0071735

the model difference between the two model is significant and the homogenous association model is a better fit compared the mutual independence model.

part f

(PI, SI), (PS, IS), (PI ,PS)

```
inc_pol_sex <- xtabs(data=montana, ~Inc+Pol+Sex)
#model
inc_pol <- as.data.frame(inc_pol)
```

```
PI.SI <- glm(Freq ~ Inc + Pol + Sex + Sex * Inc + Pol * Inc, data = inc_pol_sex, family = poisson())
summary(PI.SI)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol + Sex + Sex * Inc + Pol * Inc,
##      family = poisson(), data = inc_pol_sex)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05981  -0.46345   0.00243   0.40490   1.26709
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.07944    0.29580   7.030 2.07e-12 ***
## Inc2           0.41848    0.37910   1.104  0.26966
## Inc3           0.72335    0.37206   1.944  0.05187 .
## Pol2          -1.20397    0.46547  -2.587  0.00969 **
## Pol3          -0.35667    0.34847  -1.024  0.30604
## Sex1           0.40547    0.32275   1.256  0.20901
## Inc2:Sex1     -0.51988    0.40187  -1.294  0.19579
## Inc3:Sex1     -1.33500    0.44401  -3.007  0.00264 **
## Inc2:Pol2      1.11300    0.55476   2.006  0.04483 *
## Inc3:Pol2      0.01439    0.63482   0.023  0.98192
## Inc2:Pol3      0.47928    0.45096   1.063  0.28788
## Inc3:Pol3      0.35667    0.45649   0.781  0.43460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 46.9276  on 17  degrees of freedom
## Residual deviance:  6.5788  on  6  degrees of freedom
## AIC: 100.36
##
## Number of Fisher Scoring iterations: 4
```

```
PS.IS <- glm(Freq ~ Inc + Pol + Sex + Sex * Pol + Inc * Sex, data = inc_pol_sex, family = poisson())
summary(PS.IS)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol + Sex + Sex * Pol + Inc * Sex,
##      family = poisson(), data = inc_pol_sex)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19331  -0.67151  -0.07349   0.62310   1.41651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.6957    0.2894   5.859 4.66e-09 ***
```

```
## Inc2          0.8383      0.2992    2.802  0.00508 **
## Inc3          0.8650      0.2980    2.902  0.00370 **
## Pol2         -0.3895      0.2826   -1.378  0.16819
## Pol3          0.2296      0.2406    0.954  0.34004
## Sex1          0.7610      0.3743    2.033  0.04205 *
## Pol2:Sex1     -0.6009      0.4305   -1.396  0.16279
## Pol3:Sex1     -0.6069      0.3580   -1.695  0.09001 .
## Inc2:Sex1     -0.5199      0.4019   -1.294  0.19579
## Inc3:Sex1     -1.3350      0.4440   -3.007  0.00264 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 46.928 on 17 degrees of freedom
## Residual deviance: 10.456 on 8 degrees of freedom
## AIC: 100.23
##
## Number of Fisher Scoring iterations: 4
```

```
PI.PS <- glm(Freq ~ Inc + Pol + Sex + Inc * Pol + Pol * Inc, data = inc_pol_sex, family = poisson())
summary(PI.PS)
```

```
##
## Call:
## glm(formula = Freq ~ Inc + Pol + Sex + Inc * Pol + Pol * Inc,
##      family = poisson(), data = inc_pol_sex)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20502  -0.50231  -0.00288   0.53035   1.60335
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.41284    0.23421  10.302 < 2e-16 ***
## Inc2          0.13976    0.30574   0.457  0.64758
## Inc3          0.13976    0.30574   0.457  0.64758
## Pol2         -1.20397    0.46547  -2.587  0.00969 **
## Pol3         -0.35667    0.34847  -1.024  0.30604
## Sex1         -0.23419    0.15773  -1.485  0.13760
## Inc2:Pol2     1.11300    0.55476   2.006  0.04483 *
## Inc3:Pol2     0.01439    0.63482   0.023  0.98192
## Inc2:Pol3     0.47928    0.45096   1.063  0.28788
## Inc3:Pol3     0.35667    0.45649   0.781  0.43460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 46.928 on 17 degrees of freedom
## Residual deviance: 16.519 on 8 degrees of freedom
## AIC: 106.3
##
## Number of Fisher Scoring iterations: 5
```

(PI,SI) has the best fit. The result is in contrast with our conclusion from part a. From this analysis we concluded that political affiliation and sex are independent given the income level.

Problem 2

```
data("UCBAdmissions")
```

part a

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Gender * Admit,
##      family = poisson(), data = UCBAdmissions)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -19.7226   -7.8760    0.6475    7.1298   14.7146
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.50922     0.04106 134.188 < 2e-16 ***
## AdmitRejected      0.22013     0.03879   5.675 1.38e-08 ***
## GenderFemale     -0.76584     0.05128 -14.933 < 2e-16 ***
## DeptB            -0.46679     0.05274  -8.851 < 2e-16 ***
## DeptC            -0.01621     0.04649  -0.349 0.727355
## DeptD            -0.16384     0.04832  -3.391 0.000696 ***
## DeptE            -0.46850     0.05276  -8.879 < 2e-16 ***
## DeptF            -0.26752     0.04972  -5.380 7.44e-08 ***
## AdmitRejected:GenderFemale 0.61035     0.06389   9.553 < 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: 2650.1  on 23  degrees of freedom
## Residual deviance: 2004.2  on 15  degrees of freedom
## AIC: 2181.3
##
## Number of Fisher Scoring iterations: 5

## [1] 1748.16

##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Admit, family = poisson(),
##      data = UCBAdmissions)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.8559   -6.2383    0.0632    5.9427    8.8105
```



```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.87867    0.04260 138.005 < 2e-16 ***
## AdmitRejected    -0.59346    0.06838  -8.679 < 2e-16 ***
## GenderFemale     -0.38287    0.03027 -12.646 < 2e-16 ***
## DeptB            -0.48509    0.06608  -7.341 2.12e-13 ***
## DeptC            -0.62404    0.06906  -9.036 < 2e-16 ***
## DeptD            -0.80388    0.07336 -10.958 < 2e-16 ***
## DeptE            -1.40816    0.09201 -15.304 < 2e-16 ***
## DeptF            -2.56995    0.15298 -16.799 < 2e-16 ***
## AdmitRejected:DeptB 0.05059    0.10968   0.461   0.645
## AdmitRejected:DeptC 1.20915    0.09726  12.432 < 2e-16 ***
## AdmitRejected:DeptD 1.25833    0.10152  12.395 < 2e-16 ***
## AdmitRejected:DeptE 1.68296    0.11733  14.343 < 2e-16 ***
## AdmitRejected:DeptF 3.26911    0.16707  19.567 < 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: 2650.1  on 23  degrees of freedom
## Residual deviance: 1242.4  on 11  degrees of freedom
## AIC: 1427.4
##
## Number of Fisher Scoring iterations: 5

## [1] 1078.073

##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Gender +
##      Dept * Admit, family = poisson(), data = UCBAAdmissions)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4776  -0.4144   0.0098   0.3089   2.2321
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      6.27557    0.04248 147.744 < 2e-16 ***
## AdmitRejected    -0.59346    0.06838  -8.679 < 2e-16 ***
## GenderFemale     -2.03325    0.10233 -19.870 < 2e-16 ***
## DeptB            -0.40575    0.06770  -5.993 2.06e-09 ***
## DeptC            -1.53939    0.08305 -18.536 < 2e-16 ***
## DeptD            -1.32234    0.08159 -16.207 < 2e-16 ***
## DeptE            -2.40277    0.11014 -21.816 < 2e-16 ***
## DeptF            -3.09624    0.15756 -19.652 < 2e-16 ***
## GenderFemale:DeptB -1.07581    0.22860  -4.706 2.52e-06 ***
## GenderFemale:DeptC  2.63462    0.12343  21.345 < 2e-16 ***
## GenderFemale:DeptD  1.92709    0.12464  15.461 < 2e-16 ***
## GenderFemale:DeptE  2.75479    0.13510  20.391 < 2e-16 ***
## GenderFemale:DeptF  1.94356    0.12683  15.325 < 2e-16 ***
## AdmitRejected:DeptB 0.05059    0.10968   0.461   0.645
```

```

## AdmitRejected:DeptC  1.20915    0.09726  12.432 < 2e-16 ***
## AdmitRejected:DeptD  1.25833    0.10152  12.395 < 2e-16 ***
## AdmitRejected:DeptE  1.68296    0.11733  14.343 < 2e-16 ***
## AdmitRejected:DeptF  3.26911    0.16707  19.567 < 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: 2650.095  on 23  degrees of freedom
## Residual deviance:  21.736  on  6  degrees of freedom
## AIC: 216.8
##
## Number of Fisher Scoring iterations: 4

## [1] 19.93841

##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Admit + Gender *
##      Admit, family = poisson(), data = UCBAmissions)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.7456  -6.9026   0.0064   5.9891   9.0554
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      6.01678    0.04392  136.999 < 2e-16 ***
## AdmitRejected    -0.83007    0.07245  -11.456 < 2e-16 ***
## GenderFemale     -0.76584    0.05128  -14.933 < 2e-16 ***
## DeptB           -0.48509    0.06608   -7.341 2.12e-13 ***
## DeptC           -0.62404    0.06906   -9.036 < 2e-16 ***
## DeptD           -0.80388    0.07336  -10.958 < 2e-16 ***
## DeptE           -1.40816    0.09201  -15.304 < 2e-16 ***
## DeptF           -2.56995    0.15298  -16.799 < 2e-16 ***
## AdmitRejected:DeptB  0.05059    0.10968   0.461  0.645
## AdmitRejected:DeptC  1.20915    0.09726  12.432 < 2e-16 ***
## AdmitRejected:DeptD  1.25833    0.10152  12.395 < 2e-16 ***
## AdmitRejected:DeptE  1.68296    0.11733  14.344 < 2e-16 ***
## AdmitRejected:DeptF  3.26911    0.16707  19.567 < 2e-16 ***
## AdmitRejected:GenderFemale 0.61035    0.06389   9.553 < 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: 2650.1  on 23  degrees of freedom
## Residual deviance: 1148.9  on 10  degrees of freedom
## AIC: 1336
##
## Number of Fisher Scoring iterations: 5

## [1] 1015.707

```

Model	df	G^2	p-value	x^2	pvalue
(D, S, A)	16	2097.2	<.001	1999.6	<.001
(DS, A)	11	876.6	<.001	797.1	<.001
(D, SA)	15	2650.1	0	1748.16	0
(DA, S)	11	1242.4	0	1427.4	0
(DS, SA)	10	782.65	<.001	714.30	<.001
(DS, DA)	6	21.736	0.0013517	19.93841	0.0028402
(DA, SA)	10	1148.9	0	1015.707	0
(DS, DA, SA)	5	20.23	<.01	18.83	<.01
(DSA)	0	0	-	-	-

The (DS,DA) and (DS,DA,SA) model have the best fits. The conditional independence model is less complex which might generalize better.

part b

Model	df	G^2	p-value	H_0	Delta df	ΔG^2
(DS, DA, SA)	5	20.23	<.01	-	-	-
(DS, SA)	10	782.65	<.001	$\lambda_{ik}^{DA} = 0$	5	762.42
(DS, DA)	6	21.736	0.0013517	$\lambda_{jk}^{SA} = 0$	1	1.506
(DA, SA)	10	1148.9	0	$\lambda_{ij}^{DS} = 0$	5	1128.67

for (DS,DA): $\Delta G^2/df$ is small, we cannot reject the null hypothesis and we can conclude that the sex has no significant effect on after department is included in the model

for (DA, SA): $\Delta G^2/df$ is large and we can conclude that department and gender are independent after considering the addmition in the model.

part c

```
anova(DS.DA)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Freq
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                23    2650.10
## Admit             1    230.03
## Gender             1    162.87
## Dept              5    159.52
## Gender:Dept       5   1220.61
## Admit:Dept        5    855.32
```

The best model is (DS,DA). based on this model we concluded that there's no relationship between gender and admission when department is included in the model. The residual deviance is equal to 21.73 with

part d

```
df <- as.data.frame(UCBAdmissions)
df <- df[df$Dept != 'A',]

berk.sat <- glm(Freq ~ Admit * Gender * Dept,data = df, family = poisson())
summary(berk.sat)
```

```
##
## Call:
## glm(formula = Freq ~ Admit * Gender * Dept, family = poisson(),
##      data = df)
##
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.86647    0.05322 110.221 < 2e-16 ***
## AdmitRejected    -0.53375    0.08754  -6.097 1.08e-09 ***
## GenderFemale     -3.03325    0.24831 -12.216 < 2e-16 ***
## DeptC            -1.07898    0.10567 -10.211 < 2e-16 ***
## DeptD            -0.93921    0.10040  -9.355 < 2e-16 ***
## DeptE            -1.89618    0.14731 -12.872 < 2e-16 ***
## DeptF            -2.77543    0.21974 -12.630 < 2e-16 ***
## AdmitRejected:GenderFemale -0.22002    0.43759  -0.503  0.615
## AdmitRejected:DeptC    1.06927    0.14448   7.401 1.35e-13 ***
## AdmitRejected:DeptD    1.23771    0.13599   9.101 < 2e-16 ***
## AdmitRejected:DeptE    1.49071    0.18379   8.111 5.02e-16 ***
## AdmitRejected:DeptF    3.30349    0.23657  13.964 < 2e-16 ***
## GenderFemale:DeptC     3.55403    0.27375  12.983 < 2e-16 ***
## GenderFemale:DeptD     2.98120    0.27665  10.776 < 2e-16 ***
## GenderFemale:DeptE     3.60626    0.30193  11.944 < 2e-16 ***
## GenderFemale:DeptF     3.12027    0.38572   8.090 5.99e-16 ***
## AdmitRejected:GenderFemale:DeptC 0.34494    0.46066   0.749  0.454
## AdmitRejected:GenderFemale:DeptD 0.13804    0.46266   0.298  0.765
## AdmitRejected:GenderFemale:DeptE 0.42021    0.48123   0.873  0.383
## AdmitRejected:GenderFemale:DeptF 0.03113    0.53349   0.058  0.953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance:  1.8798e+03  on 19  degrees of freedom
## Residual deviance: -4.9516e-14  on  0  degrees of freedom
## AIC: 172.28
##
## Number of Fisher Scoring iterations: 3
```

```
#----- (DS, DA, SA) -----
homo <- glm(Freq ~ Admit + Gender + Dept + Gender * Admit + Gender * Dept + Dept * Admit, data = df, family = poisson())
summary(homo)
```

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Gender * Admit +
##      Gender * Dept + Dept * Admit, family = poisson(), data = df)
##
## Deviance Residuals:
##      5      6      7      8      9     10     11
## -0.07226  0.09469  0.33892 -0.45581  0.41912 -0.31429 -0.31658
##     12     13     14     15     16     17     18
##  0.23013 -0.41987  0.30062  0.44176 -0.31738  0.58970 -0.35213
##     19     20     21     22     23     24
## -0.42470  0.24275 -0.48624  0.12626  0.49853 -0.13223
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.87031    0.05272 111.352 < 2e-16 ***
## AdmitRejected    -0.54418    0.08584  -6.340 2.3e-10 ***
## GenderFemale     -3.12044    0.20697 -15.076 < 2e-16 ***
## DeptC            -1.12133    0.09509 -11.792 < 2e-16 ***
## DeptD            -0.90753    0.09086  -9.988 < 2e-16 ***
## DeptE            -1.98213    0.12167 -16.291 < 2e-16 ***
## DeptF            -2.67736    0.16466 -16.260 < 2e-16 ***
## AdmitRejected:GenderFemale  0.03069    0.08676   0.354  0.724
## GenderFemale:DeptC    3.70192    0.21705 17.056 < 2e-16 ***
## GenderFemale:DeptD    2.99401    0.21787 13.742 < 2e-16 ***
## GenderFemale:DeptE    3.81904    0.22496 16.976 < 2e-16 ***
## GenderFemale:DeptF    3.00203    0.22310 13.456 < 2e-16 ***
## AdmitRejected:DeptC    1.14008    0.12188   9.354 < 2e-16 ***
## AdmitRejected:DeptD    1.19456    0.11984   9.968 < 2e-16 ***
## AdmitRejected:DeptE    1.61308    0.13928 11.581 < 2e-16 ***
## AdmitRejected:DeptF    3.20527    0.17880 17.927 < 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: 1879.7761  on 19  degrees of freedom
## Residual deviance:   2.5564  on  4  degrees of freedom
## AIC: 166.84
##
## Number of Fisher Scoring iterations: 3
```

```
sum(residuals(homo, type = "pearson")^2)
```

```
## [1] 2.558182
```

```
#----- (DS.A) -----
DS.A <- glm(Freq ~ Admit + Gender + Dept + Gender * Admit, data = df, family = poisson())
summary(DS.A)
```

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Gender * Admit,
##      family = poisson(), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.4953  -5.9707  -0.0155   3.8170  18.1602
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.715747   0.053745  87.743 < 2e-16 ***
## AdmitRejected      0.542392   0.048012  11.297 < 2e-16 ***
## GenderFemale     -0.382409   0.059953  -6.378 1.79e-10 ***
## DeptC              0.450586   0.052901   8.518 < 2e-16 ***
## DeptD              0.302950   0.054514   5.557 2.74e-08 ***
## DeptE             -0.001711   0.058494  -0.029 0.976666
## DeptF              0.199271   0.055765   3.573 0.000352 ***
## AdmitRejected:GenderFemale 0.447213   0.072361   6.180 6.40e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1879.8  on 19  degrees of freedom
## Residual deviance: 1254.3  on 12  degrees of freedom
## AIC: 1402.6
##
## Number of Fisher Scoring iterations: 5
```

```
sum(residuals(DS.A, type = "pearson")^2)
```

```
## [1] 1208.264
```

```
#----- (D,SA) -----
D.SA <- glm(Freq ~ Admit + Gender + Dept + Gender * Admit, data = df, family = poisson())
summary(D.SA)
```

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Gender * Admit,
##      family = poisson(), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.4953  -5.9707  -0.0155   3.8170  18.1602
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.715747   0.053745  87.743 < 2e-16 ***
## AdmitRejected      0.542392   0.048012  11.297 < 2e-16 ***
## GenderFemale     -0.382409   0.059953  -6.378 1.79e-10 ***
## DeptC              0.450586   0.052901   8.518 < 2e-16 ***
```

```
## DeptD          0.302950    0.054514    5.557 2.74e-08 ***
## DeptE          -0.001711    0.058494   -0.029 0.976666
## DeptF           0.199271    0.055765    3.573 0.000352 ***
## AdmitRejected:GenderFemale 0.447213    0.072361    6.180 6.40e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 1879.8  on 19  degrees of freedom
## Residual deviance: 1254.3  on 12  degrees of freedom
## AIC: 1402.6
##
## Number of Fisher Scoring iterations: 5
```

```
sum(residuals(D.SA, type = "pearson")^2)
```

```
## [1] 1208.264
```

```
#----- (DA,S) -----
DA.S <- glm(Freq ~ Admit + Gender + Dept + Dept * Admit, data = df, family = poisson())
summary(DA.S)
```

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Admit, family = poisson(),
## data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -15.5520   -3.0836    0.0025    2.8876   10.3765
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.25831    0.05441  96.647 < 2e-16 ***
## AdmitRejected    -0.54286    0.08575  -6.331 2.44e-10 ***
## GenderFemale     -0.07741    0.03339  -2.318  0.0204 *
## DeptC            -0.13895    0.07621  -1.823  0.0683 .
## DeptD            -0.31879    0.08013  -3.979 6.93e-05 ***
## DeptE            -0.92307    0.09750  -9.468 < 2e-16 ***
## DeptF            -2.08486    0.15634 -13.336 < 2e-16 ***
## AdmitRejected:DeptC  1.15855    0.11017  10.516 < 2e-16 ***
## AdmitRejected:DeptD  1.20774    0.11394  10.599 < 2e-16 ***
## AdmitRejected:DeptE  1.63237    0.12824  12.729 < 2e-16 ***
## AdmitRejected:DeptF  3.21851    0.17490  18.402 < 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: 1879.78  on 19  degrees of freedom
## Residual deviance:  756.12  on  9  degrees of freedom
## AIC: 910.4
```

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

sum(residuals(DA.S, type = "pearson")^2)

## [1] 638.9159

#----- (DS,DA) -----
DS.DA <- glm(Freq ~ Admit + Gender + Dept + Dept * Gender + Dept * Admit, data = df, family = poisson())
summary(DS.DA)

##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Gender +
##      Dept * Admit, family = poisson(), data = df)
##
## Deviance Residuals:
##      5      6      7      8      9     10     11
## -0.06316  0.08273  0.29514 -0.40088  0.55733 -0.41519 -0.41820
##     12     13     14     15     16     17     18
##  0.30511 -0.30655  0.21843  0.32036 -0.23141  0.69837 -0.41419
##     19     20     21     22     23     24
## -0.49916  0.28628 -0.42032  0.10861  0.42684 -0.11382
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    5.86983    0.05272  111.35 < 2e-16 ***
## AdmitRejected -0.54286    0.08575   -6.33 2.44e-10 ***
## GenderFemale  -3.10906    0.20441  -15.21 < 2e-16 ***
## DeptC         -1.13365    0.08873  -12.78 < 2e-16 ***
## DeptD         -0.91659    0.08736  -10.49 < 2e-16 ***
## DeptE        -1.99702    0.11448  -17.45 < 2e-16 ***
## DeptF        -2.69049    0.16062  -16.75 < 2e-16 ***
## GenderFemale:DeptC  3.71043    0.21575   17.20 < 2e-16 ***
## GenderFemale:DeptD  3.00290    0.21645   13.87 < 2e-16 ***
## GenderFemale:DeptE  3.83060    0.22263   17.21 < 2e-16 ***
## GenderFemale:DeptF  3.01937    0.21771   13.87 < 2e-16 ***
## AdmitRejected:DeptC  1.15855    0.11017   10.52 < 2e-16 ***
## AdmitRejected:DeptD  1.20774    0.11394   10.60 < 2e-16 ***
## AdmitRejected:DeptE  1.63237    0.12824   12.73 < 2e-16 ***
## AdmitRejected:DeptF  3.21851    0.17490   18.40 < 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: 1879.7761  on 19  degrees of freedom
## Residual deviance:   2.6815  on  5  degrees of freedom
## AIC: 164.96
##
## Number of Fisher Scoring iterations: 3
```



```
sum(residuals(DS.DA, type = "pearson")^2)
```

```
## [1] 2.6904
```

```
#----- (DA,SA) -----
```

```
DA.SA <- glm(Freq ~ Admit + Gender + Dept + Dept * Admit + Gender * Admit, data = df, family = poisson())
summary(DA.SA)
```

```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept + Dept * Admit + Gender *
##      Admit, family = poisson(), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.859   -4.273    0.063    4.218    8.884
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.39339    0.05739  93.976 < 2e-16 ***
## AdmitRejected    -0.74883    0.09155  -8.179 2.86e-16 ***
## GenderFemale     -0.38241    0.05995  -6.378 1.79e-10 ***
## DeptC            -0.13895    0.07621  -1.823  0.0683 .
## DeptD            -0.31879    0.08013  -3.979 6.93e-05 ***
## DeptE            -0.92307    0.09750  -9.468 < 2e-16 ***
## DeptF           -2.08486    0.15634 -13.336 < 2e-16 ***
## AdmitRejected:DeptC  1.15855    0.11017  10.516 < 2e-16 ***
## AdmitRejected:DeptD  1.20774    0.11394  10.600 < 2e-16 ***
## AdmitRejected:DeptE  1.63237    0.12824  12.729 < 2e-16 ***
## AdmitRejected:DeptF  3.21851    0.17490  18.402 < 2e-16 ***
## AdmitRejected:GenderFemale 0.44721    0.07236   6.180 6.40e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 1879.78  on 19  degrees of freedom
## Residual deviance:  717.51  on  8  degrees of freedom
## AIC: 873.79
##
## Number of Fisher Scoring iterations: 5
```

```
sum(residuals(DA.SA, type = "pearson")^2)
```

```
## [1] 610.8882
```

comparing homogenous Association model and the conditional independence model (DA,SA): $\Delta G^2 = 0.0952$, with $\Delta df = 1$

the p -value is equal to .75 and the difference between the two model is not significant. The answer chosen model is the same as in part c (DS,DA). The model fits well with residual deviance 2.65 and $df = 5$. This model states that the gender and admission are independent given the department.

```
anova(DS.DA)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Freq
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                19    1879.78
## Admit           1    469.90        18    1409.87
## Gender          1      5.38        17    1404.49
## Dept            4    111.60        13    1292.90
## Gender:Dept     4    753.44          9     539.46
## Admit:Dept      4    536.78          5       2.68
```

All variables are significant based on the result of anova.

Problem 3

The X and Y are conditionally independent. All terms in the saturated model that are not in the model (WXZ, WYZ) involve X and Y, which permits XY as a conditional association.