

# Log Linear Models

## Instructions

An analysis of admissions data at UC Berkely initially suggested a gender bias, such that men were admitted at a higher rate than women. However, a more detailed analysis reveals a three-way relationship between gender, admission decision, and department. It turns out that women applied to more selective departments and in one department (least selective) the admit rate for women was actually higher than for men (the opposite of the initial result).

These data are available in R, data = UCBAAdmissions

```
library(pacman)
```

```
## Warning: package 'pacman' was built under R version 3.5.3
```

```
p_load(vcd, MASS)
```

Question: is there gender bias in this sample?

```
##Initial analysis Gender x Admit
```

```
Gender.Admit <- xtabs(Freq ~ Admit + Gender, data=UCBAAdmissions)
```

```
# Note that this data frame has an entire column for frequency data (this data frame is not setup for o
```

```
chisq.test(Gender.Admit) # requires matrix or vector, contingency table (xtabs) matches
```

```
##
```

```
## Pearson's Chi-squared test with Yates' continuity correction
```

```
##
```

```
## data: Gender.Admit
```

```
## X-squared = 91.61, df = 1, p-value < 2.2e-16
```

```
assocstats(Gender.Admit) # requires contingency table (xtabs)
```

```
##                X^2 df P(> X^2)
```

```
## Likelihood Ratio 93.449  1      0
```

```
## Pearson          92.205  1      0
```

```
##
```

```
## Phi-Coefficient   : 0.143
```

```
## Contingency Coeff.: 0.141
```

```
## Cramer's V        : 0.143
```

```
Gender.Admit
```

```
##           Gender
```

```
## Admit      Male Female
```

```
##   Admitted 1198    557
```

```
##   Rejected 1493   1278
```

```
UCBAAdmissions
```

```
## , , Dept = A
```

```
##
```

```
##           Gender
```

```
## Admit      Male Female
```

```
##   Admitted  512     89
```

```
##   Rejected  313     19
```

```
##
```

```
## , , Dept = B
##
##           Gender
## Admit      Male Female
##   Admitted  353     17
##   Rejected  207      8
##
## , , Dept = C
##
##           Gender
## Admit      Male Female
##   Admitted  120     202
##   Rejected  205     391
##
## , , Dept = D
##
##           Gender
## Admit      Male Female
##   Admitted  138     131
##   Rejected  279     244
##
## , , Dept = E
##
##           Gender
## Admit      Male Female
##   Admitted   53      94
##   Rejected  138     299
##
## , , Dept = F
##
##           Gender
## Admit      Male Female
##   Admitted   22      24
##   Rejected  351     317
```

##Log linear models This is a teaching tool that matches slide 24 of the PowerPoint

```
modell1 <- loglm( ~ Admit + Dept + Gender, data=UCBAdmissions) # requires table (this dataset is of this
summary(modell1)
```

```
## Formula:
## ~Admit + Dept + Gender
## attr("variables")
## list(Admit, Dept, Gender)
## attr("factors")
##      Admit Dept Gender
## Admit    1    0      0
## Dept     0    1      0
## Gender   0    0      1
## attr("term.labels")
## [1] "Admit" "Dept"  "Gender"
## attr("order")
## [1] 1 1 1
## attr("intercept")
## [1] 1
```

```
## attr("response")
## [1] 0
## attr(,".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##               X^2 df P(> X^2)
## Likelihood Ratio 2097.671 16      0
## Pearson          2000.328 16      0

model2 <- loglm( ~ Admit*Dept + Gender, data=UCBAdmissions)
summary(model2)
```

```
## Formula:
## ~Admit * Dept + Gender
## attr("variables")
## list(Admit, Dept, Gender)
## attr("factors")
##      Admit Dept Gender Admit:Dept
## Admit      1   0      0          1
## Dept       0   1      0          1
## Gender     0   0      1          0
## attr("term.labels")
## [1] "Admit"      "Dept"          "Gender"        "Admit:Dept"
## attr("order")
## [1] 1 1 1 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(,".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##               X^2 df P(> X^2)
## Likelihood Ratio 1242.350 11      0
## Pearson          1078.073 11      0

model3 <- loglm( ~ Admit*Gender + Dept, data=UCBAdmissions)
summary(model3)
```

```
## Formula:
## ~Admit * Gender + Dept
## attr("variables")
## list(Admit, Gender, Dept)
## attr("factors")
##      Admit Gender Dept Admit:Gender
## Admit      1      0   0          1
## Gender     0      1   0          1
## Dept       0      0   1          0
## attr("term.labels")
## [1] "Admit"      "Gender"      "Dept"        "Admit:Gender"
## attr("order")
## [1] 1 1 1 2
## attr("intercept")
```

```
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##              X^2 df P(> X^2)
## Likelihood Ratio 2004.222 15      0
## Pearson          1748.160 15      0

model4 <- loglm( ~ Admit + Dept*Gender, data = UCBAAdmissions)
summary(model4)
```

```
## Formula:
## ~Admit + Dept * Gender
## attr("variables")
## list(Admit, Dept, Gender)
## attr("factors")
##      Admit Dept Gender Dept:Gender
## Admit      1   0      0           0
## Dept       0   1      0           1
## Gender     0   0      1           1
## attr("term.labels")
## [1] "Admit"      "Dept"          "Gender"        "Dept:Gender"
## attr("order")
## [1] 1 1 1 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##              X^2 df P(> X^2)
## Likelihood Ratio 877.0564 11      0
## Pearson          797.7045 11      0
```

```
model5 <- loglm( ~ Admit*Dept + Admit*Gender, data = UCBAAdmissions)
summary(model5)
```

```
## Formula:
## ~Admit * Dept + Admit * Gender
## attr("variables")
## list(Admit, Dept, Gender)
## attr("factors")
##      Admit Dept Gender Admit:Dept Admit:Gender
## Admit      1   0      0           1           1
## Dept       0   1      0           1           0
## Gender     0   0      1           0           1
## attr("term.labels")
## [1] "Admit"      "Dept"          "Gender"        "Admit:Dept"
## [5] "Admit:Gender"
## attr("order")
```

```
## [1] 1 1 1 2 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##               X^2 df P(> X^2)
## Likelihood Ratio 1148.901 10      0
## Pearson          1015.707 10      0

model6 <- loglm( ~ Admit*Dept + Dept*Gender, data = UCBAAdmissions)
summary(model6)

## Formula:
## ~Admit * Dept + Dept * Gender
## attr("variables")
## list(Admit, Dept, Gender)
## attr("factors")
##      Admit Dept Gender Admit:Dept Dept:Gender
## Admit      1   0      0          1          0
## Dept       0   1      0          1          1
## Gender      0   0      1          0          1
## attr("term.labels")
## [1] "Admit"      "Dept"          "Gender"         "Admit:Dept"    "Dept:Gender"
## attr("order")
## [1] 1 1 1 2 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##               X^2 df    P(> X^2)
## Likelihood Ratio 21.73551  6 0.001351993
## Pearson          19.93841  6 0.002840164

model7 <- loglm( ~ Admit*Gender + Dept*Gender, data = UCBAAdmissions)
summary(model7)

## Formula:
## ~Admit * Gender + Dept * Gender
## attr("variables")
## list(Admit, Gender, Dept)
## attr("factors")
##      Admit Gender Dept Admit:Gender Gender:Dept
## Admit      1      0   0          1          0
## Gender      0      1   0          1          1
## Dept       0      0   1          0          1
## attr("term.labels")
## [1] "Admit"      "Gender"      "Dept"        "Admit:Gender"
```

```
## [5] "Gender:Dept"
## attr("order")
## [1] 1 1 1 2 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##
##              X^2 df P(> X^2)
## Likelihood Ratio 783.6070 10      0
## Pearson          715.2958 10      0

model8 <- loglm( ~ Admit*Gender + Dept*Gender + Admit*Dept, data = UCBAAdmissions)
summary(model8)
```

```
## Formula:
## ~Admit * Gender + Dept * Gender + Admit * Dept
## attr("variables")
## list(Admit, Gender, Dept)
## attr("factors")
##
##      Admit Gender Dept Admit:Gender Gender:Dept Admit:Dept
## Admit      1      0      0              1              0              1
## Gender      0      1      0              1              1              0
## Dept        0      0      1              0              1              1
## attr("term.labels")
## [1] "Admit"      "Gender"      "Dept"      "Admit:Gender"
## [5] "Gender:Dept" "Admit:Dept"
## attr("order")
## [1] 1 1 1 2 2 2
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##
##              X^2 df      P(> X^2)
## Likelihood Ratio 20.20428  5 0.001144076
## Pearson          18.82345  5 0.002073224

model9 <- loglm( ~ Admit*Gender*Dept, data = UCBAAdmissions)
summary(model9)
```

```
## Formula:
## ~Admit * Gender * Dept
## attr("variables")
## list(Admit, Gender, Dept)
## attr("factors")
##
##      Admit Gender Dept Admit:Gender Admit:Dept Gender:Dept
## Admit      1      0      0              1              1              0
## Gender      0      1      0              1              0              1
```

```
## Dept      0      0      1      0      1      1
##      Admit:Gender:Dept
## Admit      1
## Gender      1
## Dept      1
## attr("term.labels")
## [1] "Admit"      "Gender"      "Dept"
## [4] "Admit:Gender" "Admit:Dept"  "Gender:Dept"
## [7] "Admit:Gender:Dept"
## attr("order")
## [1] 1 1 1 2 2 2 3
## attr("intercept")
## [1] 1
## attr("response")
## [1] 0
## attr(".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##              X^2 df P(> X^2)
## Likelihood Ratio    0  0      1
## Pearson              0  0      1
```

## Compare models

```
stats::anova(model1, model2, model3, model4, model5, model6, model7, model8, model9)
```

## LR tests for hierarchical log-linear models

```
##
## Model 1:
## ~Admit + Dept + Gender
## Model 2:
## ~Admit * Gender + Dept
## Model 3:
## ~Admit * Dept + Gender
## Model 4:
## ~Admit + Dept * Gender
## Model 5:
## ~Admit * Dept + Admit * Gender
## Model 6:
## ~Admit * Gender + Dept * Gender
## Model 7:
## ~Admit * Dept + Dept * Gender
## Model 8:
## ~Admit * Gender + Dept * Gender + Admit * Dept
## Model 9:
## ~Admit * Gender * Dept
##
```

	Deviance	df	Delta(Dev)	Delta(df)	P(> Delta(Dev))
## Model 1	2097.67121	16			
## Model 2	2004.22181	15	93.449407	1	0.00000
## Model 3	1242.35031	11	761.871499	4	0.00000
## Model 4	877.05641	11	365.293893	0	0.00000
## Model 5	1148.90090	10	-271.844486	1	1.00000
## Model 6	783.60701	10	365.293893	0	0.00000

## Model 7	21.73551	6	761.871499	4	0.00000
## Model 8	20.20428	5	1.531226	1	0.21593
## Model 9	0.00000	0	20.204281	5	0.00114
## Saturated	0.00000	0	0.000000	0	1.00000

## Proportions broken down by department

```
prop.table(UCBAdmissions[,,"A"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.6206061 0.8240741
##   Rejected 0.3793939 0.1759259
```

```
prop.table(UCBAdmissions[,,"B"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.6303571 0.6800000
##   Rejected 0.3696429 0.3200000
```

```
prop.table(UCBAdmissions[,,"C"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.3692308 0.3406408
##   Rejected 0.6307692 0.6593592
```

```
prop.table(UCBAdmissions[,,"D"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.3309353 0.3493333
##   Rejected 0.6690647 0.6506667
```

```
prop.table(UCBAdmissions[,,"E"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.2774869 0.2391858
##   Rejected 0.7225131 0.7608142
```

```
prop.table(UCBAdmissions[,,"F"], margin=2)
```

```
##           Gender
## Admit      Male   Female
##   Admitted 0.05898123 0.07038123
##   Rejected 0.94101877 0.92961877
```

## glm approach to get expected values, deviations, and odds

this is a very good way to see how your model performs

```
mod1 <- glm(Freq ~ Admit + Gender + Dept, data = UCBAdmissions, family = poisson)
summary(mod1)
```



```
##
## Call:
## glm(formula = Freq ~ Admit + Gender + Dept, family = poisson,
##      data = UCBAAdmissions)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.170   -7.719   -1.008    4.734   17.153
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.37111    0.03964 135.498 < 2e-16 ***
## AdmitRejected  0.45674    0.03051  14.972 < 2e-16 ***
## GenderFemale  -0.38287    0.03027 -12.647 < 2e-16 ***
## DeptB         -0.46679    0.05274  -8.852 < 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 ***
## ---
## 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: 2097.7  on 16  degrees of freedom
## AIC: 2272.7
##
## Number of Fisher Scoring iterations: 5
```

```
fittedmod1 <- as.data.frame(fitted(mod1))

tab1 <- cbind(mod1$data, fittedmod1)
tab1$Dev <- tab1$Freq - tab1$fitted(mod1)
tab1
```

	Admit	Gender	Dept	Freq	fitted(mod1)	Dev
## 1	Admitted	Male	A	512	215.10146	296.898538
## 2	Rejected	Male	A	313	339.62744	-26.627437
## 3	Admitted	Female	A	89	146.67825	-57.678255
## 4	Rejected	Female	A	19	231.59285	-212.592846
## 5	Admitted	Male	B	353	134.87069	218.129308
## 6	Rejected	Male	B	207	212.94968	-5.949680
## 7	Admitted	Female	B	17	91.96868	-74.968681
## 8	Rejected	Female	B	8	145.21095	-137.210949
## 9	Admitted	Male	C	120	211.64324	-91.643240
## 10	Rejected	Male	C	205	334.16719	-129.167189
## 11	Admitted	Female	C	202	144.32008	57.679917
## 12	Rejected	Female	C	391	227.86949	163.130512
## 13	Admitted	Male	D	138	182.59417	-44.594167
## 14	Rejected	Male	D	279	288.30110	-9.301104
## 15	Admitted	Female	D	131	124.51144	6.488555
## 16	Rejected	Female	D	244	196.59328	47.406716
## 17	Admitted	Male	E	53	134.64014	-81.640144
## 18	Rejected	Male	E	138	212.58566	-74.585663

```
## 19 Admitted Female      E   94      91.81147    2.188531
## 20 Rejected Female      E  299     144.96272   154.037276
## 21 Admitted Male        F   22     164.61141  -142.611409
## 22 Rejected Male        F  351     259.90781    91.092186
## 23 Admitted Female      F   24     112.24895   -88.248954
## 24 Rejected Female      F  317     177.23182   139.768176
```

```
exp(coef(mod1))
```

```
##      (Intercept) AdmitRejected GenderFemale      DeptB      DeptC
##    215.1014624    1.5789174      0.6819026    0.6270096    0.9839228
##           DeptD           DeptE           DeptF
##      0.8488746      0.6259378      0.7652733
```

```
1/exp(coef(mod1))
```

```
##      (Intercept) AdmitRejected GenderFemale      DeptB      DeptC
##    0.004648969    0.633345363    1.466485013    1.594871791    1.016339869
##           DeptD           DeptE           DeptF
##    1.178030303    1.597602740    1.306722689
```

## Model 8

```
mod8 <- glm(Freq ~ Admit*Gender + Admit*Dept + Gender*Dept, data = UCBA admissions, family = poisson)
summary(mod8)
```

```
##
## Call:
## glm(formula = Freq ~ Admit * Gender + Admit * Dept + Gender *
##      Dept, family = poisson, data = UCBA admissions)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7
## -0.75481  0.99471  1.96454 -3.15768 -0.03402  0.04449  0.15709
##      8      9     10     11     12     13     14
## -0.22034  1.01273 -0.73839 -0.74367  0.54896  0.06760 -0.04741
##     15     16     17     18     19     20     21
## -0.06911  0.05080  1.05578 -0.61236 -0.73617  0.42678 -0.20117
##     22     23     24
##  0.05113  0.19803 -0.05370
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      6.27150    0.04271 146.855 < 2e-16 ***
## AdmitRejected    -0.58205    0.06899  -8.436 < 2e-16 ***
## GenderFemale     -1.99859    0.10593 -18.866 < 2e-16 ***
## DeptB            -0.40322    0.06784  -5.944 2.78e-09 ***
## DeptC            -1.57790    0.08949 -17.632 < 2e-16 ***
## DeptD            -1.35000    0.08526 -15.834 < 2e-16 ***
## DeptE            -2.44982    0.11755 -20.840 < 2e-16 ***
## DeptF            -3.13787    0.16174 -19.401 < 2e-16 ***
## AdmitRejected:GenderFemale -0.09987    0.08085  -1.235  0.217
## AdmitRejected:DeptB      0.04340    0.10984   0.395  0.693
## AdmitRejected:DeptC      1.26260    0.10663  11.841 < 2e-16 ***
```

```
## AdmitRejected:DeptD      1.29461    0.10582  12.234 < 2e-16 ***
## AdmitRejected:DeptE      1.73931    0.12611  13.792 < 2e-16 ***
## AdmitRejected:DeptF      3.30648    0.16998  19.452 < 2e-16 ***
## GenderFemale:DeptB      -1.07482    0.22861  -4.701 2.58e-06 ***
## GenderFemale:DeptC       2.66513    0.12609  21.137 < 2e-16 ***
## GenderFemale:DeptD       1.95832    0.12734  15.379 < 2e-16 ***
## GenderFemale:DeptE       2.79519    0.13925  20.073 < 2e-16 ***
## GenderFemale:DeptF       2.00232    0.13571  14.754 < 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: 20.204 on 5 degrees of freedom
```

```
## AIC: 217.26
```

```
##
```

```
## Number of Fisher Scoring iterations: 4
```

```
fittedmod8 <- as.data.frame(fitted(mod8))
```

```
tab8 <- cbind(mod8$data, fittedmod8)
```

```
tab8$Dev <- tab8$Freq - tab8$fitted(mod8)
```

```
tab8
```

```
##      Admit Gender Dept Freq fitted(mod8)      Dev
## 1  Admitted   Male   A   512   529.269919 -17.2699189
## 2  Rejected   Male   A   313   295.730081  17.2699189
## 3  Admitted Female   A    89    71.730081  17.2699189
## 4  Rejected Female   A    19    36.269919 -17.2699189
## 5  Admitted   Male   B   353   353.639509 -0.6395092
## 6  Rejected   Male   B   207   206.360491  0.6395092
## 7  Admitted Female   B    17    16.360491  0.6395092
## 8  Rejected Female   B     8     8.639509 -0.6395092
## 9  Admitted   Male   C   120   109.245276  10.7547236
## 10 Rejected   Male   C   205   215.754724 -10.7547236
## 11 Admitted Female   C   202   212.754724 -10.7547236
## 12 Rejected Female   C   391   380.245276  10.7547236
## 13 Admitted   Male   D   138   137.207390  0.7926100
## 14 Rejected   Male   D   279   279.792610 -0.7926100
## 15 Admitted Female   D   131   131.792610 -0.7926100
## 16 Rejected Female   D   244   243.207390  0.7926100
## 17 Admitted   Male   E    53    45.680810  7.3191902
## 18 Rejected   Male   E   138   145.319190 -7.3191902
## 19 Admitted Female   E    94   101.319190 -7.3191902
## 20 Rejected Female   E   299   291.680810  7.3191902
## 21 Admitted   Male   F    22    22.957096 -0.9570957
## 22 Rejected   Male   F   351   350.042904  0.9570957
## 23 Admitted Female   F    24    23.042904  0.9570957
## 24 Rejected Female   F   317   317.957096 -0.9570957
```

```
exp(coef(mod8))
```

```
##      (Intercept)      AdmitRejected
## 529.26991890      0.55875097
##      GenderFemale      DeptB
```

```
##          0.13552646          0.66816476
##          DeptC          DeptD
##          0.20640749          0.25923897
##          DeptE          DeptF
##          0.08630910          0.04337502
## AdmitRejected:GenderFemale AdmitRejected:DeptB
##          0.90495497          1.04435339
## AdmitRejected:DeptC AdmitRejected:DeptD
##          3.53459252          3.64955948
## AdmitRejected:DeptE AdmitRejected:DeptF
##          5.69338935          27.28890081
## GenderFemale:DeptB GenderFemale:DeptC
##          0.34135906          14.36985656
## GenderFemale:DeptD GenderFemale:DeptE
##          7.08744082          16.36567060
## GenderFemale:DeptF
##          7.40621236
```

```
1/exp(coef(mod8))
```

```
##          (Intercept)          AdmitRejected
##          0.001889395          1.789706062
## GenderFemale          DeptB
##          7.378632657          1.496636844
##          DeptC          DeptD
##          4.844785389          3.857444697
##          DeptE          DeptF
##          11.586263936          23.054742009
## AdmitRejected:GenderFemale AdmitRejected:DeptB
##          1.105027353          0.957530283
## AdmitRejected:DeptC AdmitRejected:DeptD
##          0.282918043          0.274005673
## AdmitRejected:DeptE AdmitRejected:DeptF
##          0.175642300          0.036644935
## GenderFemale:DeptB GenderFemale:DeptC
##          2.929466675          0.069590117
## GenderFemale:DeptD GenderFemale:DeptE
##          0.141094653          0.061103515
## GenderFemale:DeptF
##          0.135021783
```