

Homework 5

Shun-Lung Chang, Dilip Hiremath

```
library(vcd)
library(vcdExtra)
```

1. (2 points) Construct the two-way table for gender and whether admitted. Find the odds ratio for admission of males vs. females and interpret. For which gender is the probability of admission higher?

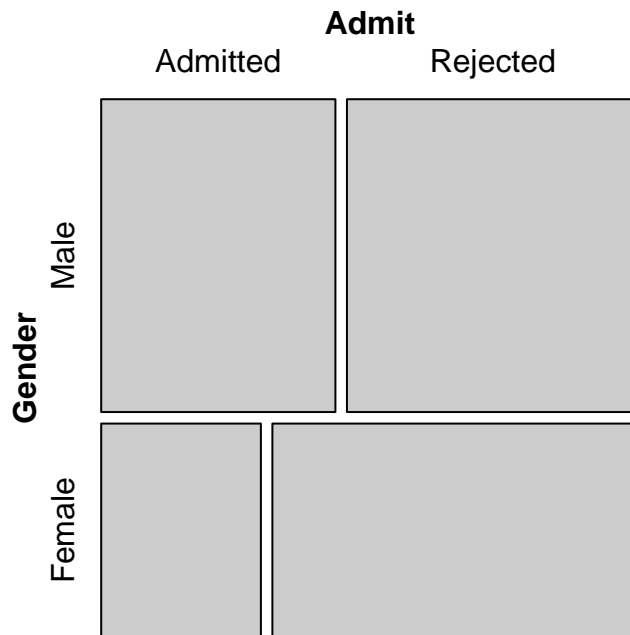
```
UCBAdmissions_m <- margin.table(UCBAdmissions, c(2, 1))
oddsratio(UCBAdmissions_m, log = FALSE)
```

odds ratios for Gender and Admit

```
[1] 1.84108
```

2. Draw a mosaic plot for the data aggregated over departments.

```
mosaic(UCBAdmissions_m)
```



3. (3 points) Fit a saturated log-linear model to the frequencies using A and G as predictors. Calculate the odds ratio for admission of males vs. females. Display the predicted frequencies in a table and compare them to the observed frequencies displayed in Question 1.

```
mod_1 <- glm(Freq ~ Admit * Gender, data = UCBAAdmissions, family = 'poisson')
summary(mod_1)
```

Call:

```
glm(formula = Freq ~ Admit * Gender, family = "poisson", data = UCBAAdmissions)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-18.9074	-7.8908	-0.1399	5.6871	18.4285

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.29665	0.02889	183.329	< 2e-16
AdmitRejected	0.22013	0.03879	5.675	0.0000000138
GenderFemale	-0.76584	0.05128	-14.933	< 2e-16
AdmitRejected:GenderFemale	0.61035	0.06389	9.553	< 2e-16

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2650.1 on 23 degrees of freedom
 Residual deviance: 2163.7 on 20 degrees of freedom
 AIC: 2330.8

Number of Fisher Scoring iterations: 5

```
exp(unname(mod_1$coefficients[4]))
```

```
[1] 1.84108
```

```
# transform the fitted value so that it has same format as UCBAAdmissions
```

```
tran_preds <- function(mod) {
  f <- mod$fitted.values
  dim(f) <- dim(UCBAAdmissions)
  dimnames(f) <- dimnames(UCBAAdmissions)
  return(f)
}
f <- tran_preds(mod_1)

mod_table <- ftable(f,
  row.vars = c("Admit", "Dept"),
  col.vars = "Gender")
mod_table
```

	Gender	Male	Female
Admit Dept			
Admitted A		199.66667	92.83333
B		199.66667	92.83333
C		199.66667	92.83333
D		199.66667	92.83333

	E	199.66667	92.83333
	F	199.66667	92.83333
Rejected	A	248.83333	213.00000
	B	248.83333	213.00000
	C	248.83333	213.00000
	D	248.83333	213.00000
	E	248.83333	213.00000
	F	248.83333	213.00000

```
obs_table <- ftable(UCBAdmissions,
  row.vars = c("Admit", "Dept"),
  col.vars = "Gender")
obs_table
```

		Gender Male Female	
Admit	Dept		
Admitted	A	512	89
	B	353	17
	C	120	202
	D	138	131
	E	53	94
	F	22	24
Rejected	A	313	19
	B	207	8
	C	205	391
	D	279	244
	E	138	299
	F	351	317

4. (2 points) Fit an independence log-linear model to the frequencies using A, D and G as predictors. Display the predicted frequencies in a table and compare them to the observed frequencies. Calculate the estimated odds ratios for admission of males vs. females for each department.

```
mod_2 <- glm(Freq ~ Admit + Gender + Dept, data = UCBAdmissions,
  family = 'poisson')
summary(mod_2)
```

Call:

```
glm(formula = Freq ~ Admit + Gender + Dept, family = "poisson",
  data = UCBAdmissions)
```

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	0.0000000744

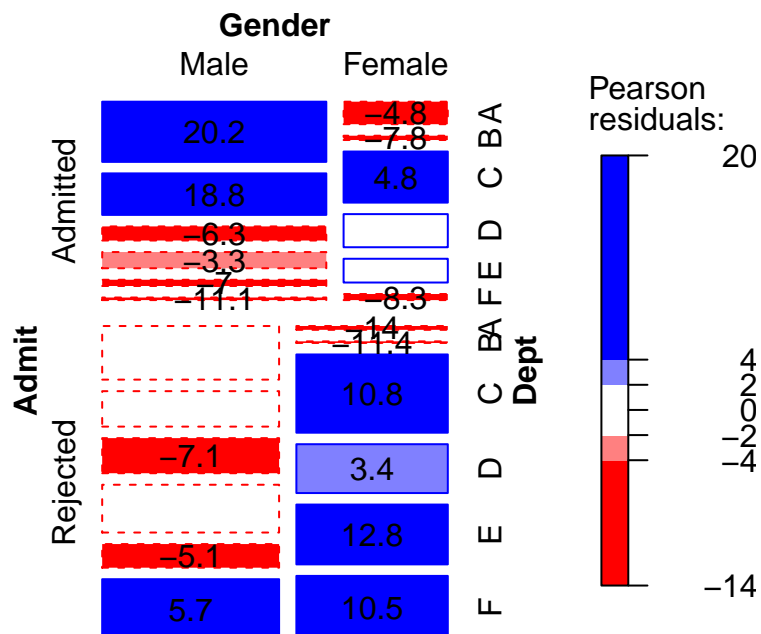
(Dispersion parameter for poisson family taken to be 1)

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

5. Draw mosaic plots of admission versus gender for each department separately.

```
mosaic.glm(mod_2, gp = shading_Friendly, labeling = labeling_residuals, formula = ~ Admit + Gender + Dept)
```



6. (3 points) Fit a log-linear model to the frequencies using A, D and G as predictors that includes all two-way interactions but not the three-way interaction. Display the predicted frequencies in a table and compare them to the predicted frequencies displayed in Question 4 as well as the observed frequencies. Calculate the estimated odds ratio for admission of males vs. females in this model.
7. (2 points) Starting with the model in Question 6, use the stepwise method with the backward/forward option and BIC as criterion. Does this result in a simpler model? Interpret this model in plain English.
8. (2 points) Calculate the saturated model using all three predictors and compare this model to the one obtained in Question 7 using AIC, BIC and the deviance criterion. Give a verbal summary of your findings.
9. To run a logistic regression model with aggregated data it is best to create a data frame that comprises two frequency variables: one for the numbers of admitted students, and one for the numbers of rejected students.

```
UCBAdmissions <- as.data.frame(UCBAdmissions)
attach(UCBAdmissions)
UCBAdmit <- cbind(UCBAdmissions[Admit=="Rejected",-1],
  A=UCBAdmissions[Admit=="Admitted", "Freq"])
names(UCBAdmit)[3] <- "R"
detach()
```

Treating `cbind(A,R)` as response and D and G as qualitative predictors, fit the logit model having main effects only.

- (a) (1.5 points) Report the prediction equation. Interpret the coefficients of D,G and the interaction in this equation.
- (b) (half a point) To which log-linear model is this model equivalent?