LINEAR MODELS AND GLM: QUIZ 2 SOLUTIONS

Question 1: For the `UCBAdmissions` dataset in R, convert it into a data frame using `as.data.frame.table()`, and

(a) Fit a suitable Poisson GLM model using the R function `glm()`
(b) Fit a suitable log-linear model with the `loglm()` function in the MASS package, which is a formula-based interface (similar to that of `glm()`) to `loglin()`.

In your answer, you only need to provide the R code you used (please use a fixed-width font).

Solution. The models fit are full interaction (saturated) models.

```
> library(MASS)
> UCBdf <- as.data.frame(as.table(UCBAdmissions))
> fm.poisson <- glm(Freq ~ Dept * Admit * Gender, data = UCBdf, family = poisson())
> fm.loglm <- loglm(Freq ~ Dept * Admit * Gender, data = UCBdf)
```

In hindsight, I should have asked you to also fit a Binomial model, as product Binomial is the more natural sampling scheme for this data. Specifically, we can think of all Department-Gender combinations as the populations of interest we condition on, and Admission as success (Rejection as failure). This needs the data to be in a slightly different format.

```
> UCBadmit <- as.data.frame(as.table(UCBAdmissions["Admitted",,,drop = FALSE]))
> UCBreject <- as.data.frame(as.table(UCBAdmissions["Rejected",,,drop = FALSE]))
> UCBadmit$Admitted <- UCBadmit$Freq; UCBadmit$Freq <- NULL; UCBadmit$Admit <- NULL
> UCBreject$Rejected <- UCBreject$Freq; UCBreject$Freq <- NULL; UCBreject$Admit <- NULL
> UCBmerged <- merge(UCBadmit, UCBreject)
> str(UCBmerged)
```

```
'data.frame':        12 obs. of  4 variables:
 $ Gender  : Factor w/ 2 levels "Male","Female": 2 2 2 2 2 2 1 1 1 1 ...
 $ Dept    : Factor w/ 6 levels "A","B","C","D",..: 1 2 3 4 5 6 1 2 3 4 ...
 $ Admitted: num  89 17 202 131 94 24 512 353 120 138 ...
 $ Rejected: num  19 8 391 244 299 317 313 207 205 279 ...
```

```
> fm.bin <- glm(cbind(Admitted, Rejected) ~ Dept * Gender,
                data = UCBmerged, family = binomial())
```

Question 2:

(a) For both approaches above, formulate and test the hypothesis that there is no gender bias in the rates of admission. Compare the results for the two approaches.
(b) If there is evidence for a gender bias, what is the direction of the bias?

Solution. The null model is obvious in the Binomial GLM case.

```
> fm0.bin <- glm(cbind(Admitted, Rejected) ~ Dept, data = UCBmerged, family = binomial())
> anova(fm0.bin, fm.bin, test = "Chisq")

Analysis of Deviance Table

Model 1: cbind(Admitted, Rejected) ~ Dept
Model 2: cbind(Admitted, Rejected) ~ Dept * Gender
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1         6     21.735
2         0      0.000  6   21.735  0.001352
```

For the Poisson and log-linear model cases, the null model is not as trivial, but a little thought should make it clear that the only terms that need to be dropped are those involving any `Gender:Admit` interactions. This leads to

```
> fm0.poisson <- glm(Freq ~ Dept * Admit + Dept * Gender, data = UCBdf, family = poisson())
> fm0.loglm <- loglm(Freq ~ Dept * Admit + Dept * Gender, data = UCBdf)
> anova(fm0.poisson, fm.poisson, test = "Chisq")

Analysis of Deviance Table

Model 1: Freq ~ Dept * Admit + Dept * Gender
Model 2: Freq ~ Dept * Admit * Gender
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1         6     21.735
2         0      0.000  6   21.735  0.001352

> anova(fm0.loglm)

Call:
loglm(formula = Freq ~ Dept * Admit + Dept * Gender, data = UCBdf)

Statistics:
                      X^2 df    P(> X^2)
Likelihood Ratio 21.73551  6 0.001351993
Pearson          19.93841  6 0.002840164
```

Naturally, all three approaches give identical results.

There is some evidence for gender bias. To isolate the direction of the bias, we can look at the signs of the extra terms in the full model. This is most conveniently extracted from the `loglm()` fit.

```
> coef.fm <- coef(fm.loglm)
> names(coef.fm)

[1] "(Intercept)"       "Dept"              "Admit"
[4] "Gender"            "Dept.Admit"        "Dept.Gender"
[7] "Admit.Gender"      "Dept.Admit.Gender"
```

```
> coef.fm[c("Admit.Gender", "Dept.Admit.Gender")]

$Admit.Gender
          Gender
Admit            Male     Female
  Admitted -0.0507447  0.0507447
  Rejected  0.0507447 -0.0507447

$Dept.Admit.Gender
, , Gender = Male

    Admit
Dept     Admitted     Rejected
   A -0.212274286  0.212274286
   B -0.004260932  0.004260932
   C  0.081975109 -0.081975109
   D  0.030247904 -0.030247904
   E  0.100791458 -0.100791458
   F  0.003520746 -0.003520746

, , Gender = Female

    Admit
Dept     Admitted     Rejected
   A  0.212274286 -0.212274286
   B  0.004260932 -0.004260932
   C -0.081975109  0.081975109
   D -0.030247904  0.030247904
   E -0.100791458  0.100791458
   F -0.003520746  0.003520746
```

The overall effect of gender on admission rates is that females have a slightly higher rate of admission. Broken up by department, the additional effect of gender on admission rates is in the same direction for all but two departments (A and B). Combining the two effects, only department A shows higher admission rates for males. In other words, the direction of the actual gender bias is the opposite of what we would have expected from the data from all departments combined.