Homework 4

Gabriel Lahman

2/7/2019

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

Question 1

```
MLE_fun_full = function(counts) {
  MLE = counts/sum(counts)
  log_lik = (counts[1,1]*log(MLE[1,1])) + (counts[1,2]*log(MLE[1,2])) + (counts[2,1]*
log(MLE[2,1])) + (counts[2,2]*log(MLE[2,2]))

  out = list(
    MLE = MLE,
    log_lik = log_lik
  )

  return (out)
}
```

Question 2

```
MLE_fun_ind = function(counts){
  row1_tot = sum(counts[1,])
  row2_tot = sum(counts[2,])
  col1_tot = sum(counts[,1])
  col2_tot = sum(counts[,2])
  row1_prob = row1_tot/(row1_tot + row2_tot)
  row2_prob = 1 - row1_prob
  col1_prob = col1_tot/(col1_tot + col2_tot)
  col2_prob = 1 - col1_prob
  MLE = matrix(nrow = 2,
               ncol = 2,
               dimnames = dimnames(counts)
               )
  MLE[1,1] = row1_prob*col1_prob
  MLE[1,2] = row1_prob*col2_prob
  MLE[2,1] = row2_prob*col1_prob
  MLE[2,2] = row2_prob*col2_prob
  log_lik = row1_tot*log(row1_prob) + row2_tot*log(row2_prob) + col1_tot*log(col1_pro
b) + col2_tot*log(col2_prob)

  out = list(
    MLE = MLE,
    log_lik = log_lik
  )

  return (out)
}
```

Question 3

```
UCB_admit_gender_tot = apply(UCBAdmissions,c("Admit","Gender"),sum)
full_model = MLE_fun_full(UCB_admit_gender_tot)
ind_model = MLE_fun_ind(UCB_admit_gender_tot)
lr_stat = -2*(ind_model$log_lik - full_model$log_lik)
p_lr = 1- pchisq(lr_stat, 1)
```

```
## [1] "Full Model"
```

```
## $MLE
##           Gender
## Admit           Male    Female
##   Admitted 0.2646929 0.1230667
##   Rejected 0.3298719 0.2823685
##
## $log_lik
## [1] -6031.193
```

```
## [1] "Ind. Model"
```

```
## $MLE
##           Gender
## Admit            Male    Female
##   Admitted 0.2305482 0.1572114
##   Rejected 0.3640165 0.2482238
##
## $log_lik
## [1] -6077.917
```

```
## [1] "Likelihood Stat.= 93.449 P-Value= 0"
```

According to the likelihood ratio statistic and the corresponding p-value, it appears as if the full model fits the data better. Since this shows support the gender and admission chances are not indpenendent random variables, there looks like may be some gender-based bias between admission rates.

Question 4

```
full_model_by_dept = apply(UCBAdmissions, 3, MLE_fun_full)
ind_model_by_dept = apply(UCBAdmissions, 3, MLE_fun_ind)

log_lik_helper = function(l1, l2){
  return (-2*(l1$log_lik - l2$log_lik))
}

lr_by_dept = mapply(log_lik_helper,
                    l1=ind_model_by_dept,
                    l2=full_model_by_dept
)

p_by_dept = lapply(lr_by_dept, (function (x) 1-pchisq(x,1)))
```

```
## [1] "Likelihood Ratio by Dept."
```

```
##           A           B           C           D           E           F
## 19.0540099  0.2586429  0.7509844  0.2978665  0.9903864  0.3836167
```

```
## [1] "P-Value by Dept."
```

```
## $A
## [1] 1.270705e-05
##
## $B
## [1] 0.611054
##
## $C
## [1] 0.3861648
##
## $D
## [1] 0.585223
##
## $E
## [1] 0.319648
##
## $F
## [1] 0.535674
```

This shows a very different conclusion than using the aggregated model. For every deparment other than A, the full model and independence model fit equally as well, saying that each gender is most likely admitted at similar rates. Only department A seems to have any gender based bias.