

Assignment4

Kenichi Obata

March 20, 2020

1. HelpDesk

Our BASIC model is Multinomial ($n=120$, p_1 , p_2 , p_3 , p_4).

```
freq <- c(52 , 38 , 21 , 9)
sum(freq) #which is 120
```

```
## [1] 120
```

```
# define log-likelihood function
#ell=estimated log likelihood
ell <- function(p, freq){
  # Multinomial log-likelihood
  # freq = frequencies; p = probabilities
  sum(freq * log(p))
}

#define LRS
LRS <- function(p0, phat, freq){
  #Likelihood ratio statistic, phat=MLE, p0=H0 values
  2*(ell(phat, freq) - ell(p0, freq))
}

#null hypothesis are following
p0 <- c(0.4 , 0.25 , 0.25 , 0.1)
#phat is MLE
phat <- freq/sum(freq)
dobs <- LRS(p0, phat, freq) #LRS observed
dobs
```

```
## [1] 6.131364
```

```
1 - pchisq(dobs, 3) #p-value k=3, q=0
```

```
## [1] 0.1053911
```

```
#expected frequency data
expected_data <- 120*p0
expected_data
```

```
## [1] 48 30 30 12
```

```
#Goodness of Fitting test
GOF <- sum((freq - expected_data)^2 / expected_data)
GOF
```

```
## [1] 5.916667
```

```
1 - pchisq(GOF,3)
```

```
## [1] 0.1157357
```

p-value is 0.1053911 which is higher than 0.1, so there is no evidence against null hypothesis. In addition, as Goodness of Fitting test shows, we CANNOT say the expected data (python=48, java=30, R=30, C#=12) display poor agreement with the observed data.

2. LNG

Our BASIC model has 3 independent Binomials($n=400$, p_1), ($n=350$, p_2), ($n=350$, p_3)

```
in_favour<-c(198,140,133)
total<-c(400,350,350)

ell<-function(y, n, p){
  # Binomial log-likelihood
  # y = number of in_favour; n=size; p=probabilities
  sum(y*log(p) + (n-y)*log(1-p))
}

LRS<-function(phat, p0, y, n){
  #Likelihood ratio statistic, phat=MLE, p0=H0 values
  # y = number of in_favour; n=size
  2*(ell(y, n, phat) - ell(y, n, p0))
}

phat<-in_favour/total #MLE under BASIC model
p0<-sum(in_favour)/sum(total) #MLE under H0
p0
```

```
## [1] 0.4281818
```

```
dobs<-LRS(phat, p0, in_favour, total) #observed LRS; k=4, q=1
dobs
```

```
## [1] 11.71976
```

```
1 - pchisq(dobs, 3-1) #p-value
```

```
## [1] 0.002851579
```

```
#expected data under H0
expected_data <- total*p0
expected_data
```

```
## [1] 171.2727 149.8636 149.8636
```

```
#Goodness of Fitting test
GOF <- sum((in_favour - expected_data)^2 / expected_data)
GOF
```

```
## [1] 6.717622
```

```
1 - pchisq(GOF,3-1)
```

```
## [1] 0.03477658
```

p-value is 0.002851579 which is less than 0.01 so we have very strong evidence against null hypothesis which assumes that probability of in favour toward being considered as possible LNG location is equal among the three locations.

Expected data under H0 is (A=171.2727, B=149.8636, C=149.8636) which display poor agreement with the observed data(A=198,B=140,C=133). Goodness of Fitting test also shows that we have evidence against expected data under null hypothesis.

3. Test of Independence - cannabis/political orientation

Our BASIC model is Multinomial(n=1350, p11, p12, p13, p21, p22, p23, p31, p32, p33).

```
freq <- matrix(c(479, 173, 119, 214, 48, 15,172,45,85), nrow=3, byrow=TRUE)
freq
```

```
##      [,1] [,2] [,3]
## [1,]  479  173  119
## [2,]  214   48   15
## [3,]  172   45   85
```

```
sum(freq)
```

```
## [1] 1350
```

```
ell <- function(p, freq){
  # Multinomial log-likelihood
  # freq = frequencies; p = probabilities
  sum(freq * log(p))
}

LRS <- function(p0, phat, freq){
  #Likelihood ratio statistic, phat=MLE, p=H0 values
  # freq=frequencies
```

```
  2*(ell(phat, freq) - ell(p0, freq))
}
```

```
rsum <- rowSums(freq)
csum <- colSums(freq)
rsum
```

```
## [1] 771 277 302
```

```
csum
```

```
## [1] 865 266 219
```

```
eij <- outer(rsum, csum) / sum(freq) # this is to calculate expected freq
eij                                     #  $e_{ij} = r_i * c_j / n$ 
```

```
##           [,1]      [,2]      [,3]
## [1,] 494.0111 151.91556 125.07333
## [2,] 177.4852  54.57926  44.93556
## [3,] 193.5037  59.50519  48.99111
```

```
# estimated probs under H0 are eij/sum(freq)
```

```
p0 <- eij / sum(freq)
phat <- freq / sum(freq)
```

```
dobs <- LRS(p0, phat, freq)
dobs
```

```
## [1] 66.38949
```

```
#pchisq gives us left side of the density so use 1 - pchisq(dobs,4) to calculate lower tail.
1 - pchisq(dobs, 4) #pvalue, df=(#rows - 1)(#cols - 1)
```

```
## [1] 1.311173e-13
```

```
#Finally, let's use the freq and eij and test PearsonGOF.
```

```
GOF <- sum((freq - eij)^2 / eij)
GOF
```

```
## [1] 64.31788
```

```
1 - pchisq(GOF,4)
```

```
## [1] 3.58269e-13
```

p-value 1.311173e-13 is less than 0.01 so we have very strong evidence against the null hypothesis which assumes that student political orientation is independent of their cannabis use. In addition, as Goodness of Fitting test shows, the expected data (494.0111, 151.91556, 125.07333, 177.4852, 54.57926, 44.93556, 193.5037, 59.50519, 48.99111) by row display poor agreement with the observed frequency. We conclude that we have very strong evidence against the null hypothesis.