hw10

Enguang Fan

4/22/2020

```r
library(poLCA)
```

```
## Warning: package 'poLCA' was built under R version 3.6.3
```

```
## Loading required package: scatterplot3d
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 3.6.2
```

```r
data("carcinoma")
carcinoma
```

```
##    A B C D E F G
## 1  1 1 1 1 1 1 1
## 2  1 1 1 1 1 1 1
## 3  1 1 1 1 1 1 1
## 4  1 1 1 1 1 1 1
## 5  1 1 1 1 1 1 1
## 6  1 1 1 1 1 1 1
## 7  1 1 1 1 1 1 1
## 8  1 1 1 1 1 1 1
## 9  1 1 1 1 1 1 1
## 10 1 1 1 1 1 1 1
## 11 1 1 1 1 1 1 1
## 12 1 1 1 1 1 1 1
## 13 1 1 1 1 1 1 1
## 14 1 1 1 1 1 1 1
## 15 1 1 1 1 1 1 1
## 16 1 1 1 1 1 1 1
## 17 1 1 1 1 1 1 1
## 18 1 1 1 1 1 1 1
## 19 1 1 1 1 1 1 1
## 20 1 1 1 1 1 1 1
## 21 1 1 1 1 1 1 1
## 22 1 1 1 1 1 1 1
## 23 1 1 1 1 1 1 1
## 24 1 1 1 1 1 1 1
## 25 1 1 1 1 1 1 1
## 26 1 1 1 1 1 1 1
```

```
## 27   1 1 1 1 1 1 1
## 28   1 1 1 1 1 1 1
## 29   1 1 1 1 1 1 1
## 30   1 1 1 1 1 1 1
## 31   1 1 1 1 1 1 1
## 32   1 1 1 1 1 1 1
## 33   1 1 1 1 1 1 1
## 34   1 1 1 1 1 1 1
## 35   1 1 1 1 2 1 1
## 36   1 1 1 1 2 1 1
## 37   1 2 1 1 1 1 1
## 38   1 2 1 1 1 1 1
## 39   1 2 1 1 1 1 1
## 40   1 2 1 1 1 1 1
## 41   1 2 1 1 1 1 1
## 42   1 2 1 1 1 1 1
## 43   1 2 1 1 1 1 2
## 44   1 2 1 1 2 1 1
## 45   1 2 1 1 2 1 1
## 46   1 2 1 1 2 1 1
## 47   1 2 1 1 2 1 1
## 48   1 2 1 1 2 1 2
## 49   1 2 1 1 2 1 2
## 50   1 2 1 1 2 1 2
## 51   1 2 1 1 2 1 2
## 52   1 2 1 1 2 1 2
## 53   2 1 1 1 1 1 1
## 54   2 1 1 1 1 1 1
## 55   2 1 2 1 2 1 2
## 56   2 2 1 1 1 1 1
## 57   2 2 1 1 1 1 1
## 58   2 2 1 1 1 1 2
## 59   2 2 1 1 2 1 1
## 60   2 2 1 1 2 1 1
## 61   2 2 1 1 2 1 2
## 62   2 2 1 1 2 1 2
## 63   2 2 1 1 2 1 2
## 64   2 2 1 1 2 1 2
## 65   2 2 1 1 2 1 2
## 66   2 2 1 1 2 1 2
## 67   2 2 1 1 2 1 2
## 68   2 2 1 1 2 2 2
## 69   2 2 1 2 1 1 2
## 70   2 2 1 2 2 1 2
## 71   2 2 1 2 2 1 2
## 72   2 2 1 2 2 2 2
## 73   2 2 1 2 2 2 2
## 74   2 2 1 2 2 2 2
## 75   2 2 2 1 2 1 2
## 76   2 2 2 1 2 1 2
## 77   2 2 2 1 2 1 2
## 78   2 2 2 1 2 1 2
## 79   2 2 2 1 2 1 2
## 80   2 2 2 1 2 1 2
```

```
## 81  2 2 2 1 2 1 2
## 82  2 2 2 1 2 1 2
## 83  2 2 2 1 2 1 2
## 84  2 2 2 1 2 1 2
## 85  2 2 2 1 2 1 2
## 86  2 2 2 1 2 1 2
## 87  2 2 2 1 2 1 2
## 88  2 2 2 1 2 2 2
## 89  2 2 2 1 2 2 2
## 90  2 2 2 1 2 2 2
## 91  2 2 2 1 2 2 2
## 92  2 2 2 1 2 2 2
## 93  2 2 2 2 2 1 2
## 94  2 2 2 2 2 1 2
## 95  2 2 2 2 2 1 2
## 96  2 2 2 2 2 1 2
## 97  2 2 2 2 2 1 2
## 98  2 2 2 2 2 1 2
## 99  2 2 2 2 2 1 2
## 100 2 2 2 2 2 1 2
## 101 2 2 2 2 2 1 2
## 102 2 2 2 2 2 1 2
## 103 2 2 2 2 2 2 2
## 104 2 2 2 2 2 2 2
## 105 2 2 2 2 2 2 2
## 106 2 2 2 2 2 2 2
## 107 2 2 2 2 2 2 2
## 108 2 2 2 2 2 2 2
## 109 2 2 2 2 2 2 2
## 110 2 2 2 2 2 2 2
## 111 2 2 2 2 2 2 2
## 112 2 2 2 2 2 2 2
## 113 2 2 2 2 2 2 2
## 114 2 2 2 2 2 2 2
## 115 2 2 2 2 2 2 2
## 116 2 2 2 2 2 2 2
## 117 2 2 2 2 2 2 2
## 118 2 2 2 2 2 2 2
```

```r
# 1 for no   2 for yes
```

```r
dim(carcinoma)
```

```
## [1] 118   7
```

```r
freq = rep(0,7)
```

```r
for (i in 1:7) {
  df = carcinoma[,i]
  df = df-1
  freq[i] = sum(df)/118
}
```
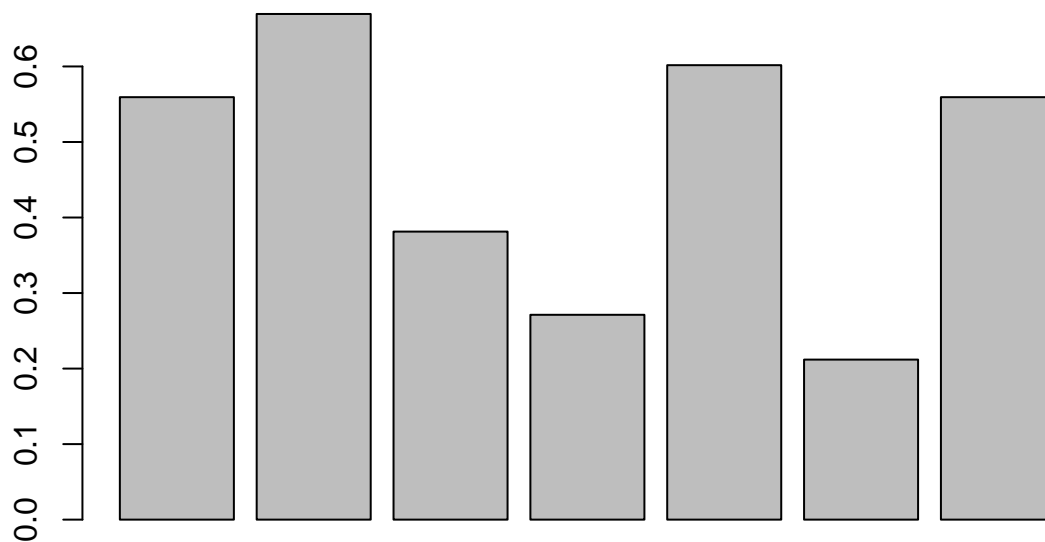
```
freq
```

```
## [1] 0.5593220 0.6694915 0.3813559 0.2711864 0.6016949 0.2118644 0.5593220
```

```
freqmtx =as.data.frame(freq)
row.names(freqmtx) = c('a','b','c','d','e','f','g')
freqmtx
```

```
##          freq
## a 0.5593220
## b 0.6694915
## c 0.3813559
## d 0.2711864
## e 0.6016949
## f 0.2118644
## g 0.5593220
```

```
barplot(freqmtx[,1])
```


Based on the bar plot, pathologist A B E and G give quite similar number of patients whereas C , D and F give smaller number, I'd say It seems to be a normal distribution ## 2

```
f=cbind(A,B,C,D,E,F,G)~1
twoclass=poLCA(f,carcinoma,nclass=2,maxiter=1000)
```

4

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $A
##           Pr(1)  Pr(2)
## class 1:  0.0000 1.0000
## class 2:  0.8835 0.1165
##
## $B
##           Pr(1)  Pr(2)
## class 1:  0.0169 0.9831
## class 2:  0.6456 0.3544
##
## $C
##           Pr(1)  Pr(2)
## class 1:  0.2391 0.7609
## class 2:  1.0000 0.0000
##
## $D
##           Pr(1)  Pr(2)
## class 1:  0.4589 0.5411
## class 2:  1.0000 0.0000
##
## $E
##           Pr(1)  Pr(2)
## class 1:  0.0214 0.9786
## class 2:  0.7771 0.2229
##
## $F
##           Pr(1)  Pr(2)
## class 1:  0.5773 0.4227
## class 2:  1.0000 0.0000
##
## $G
##           Pr(1)  Pr(2)
## class 1:  0.0000 1.0000
## class 2:  0.8835 0.1165
##
## Estimated class population shares
##  0.5012 0.4988
##
## Predicted class memberships (by modal posterior prob.)
##  0.5 0.5
##
## =========================================================
## Fit for 2 latent classes:
## =========================================================
## number of observations: 118
## number of estimated parameters: 15
## residual degrees of freedom: 103
## maximum log-likelihood: -317.2568
##
## AIC(2): 664.5137
## BIC(2): 706.0739
```

```
## G^2(2): 62.36543 (Likelihood ratio/deviance statistic)
## X^2(2): 92.64814 (Chi-square goodness of fit)
##
```

Obviously, these two classes should be interpred as have or does not have cancer. class 1 has a quite large pi value for $\Pr(2)$ and class 2 has large value for $\Pr(1)$ since $\Pr(2)$ is predicted as has cancer and $\Pr(1)$ is predicted to be healthy, we would conclude that class 1 is classified as has cancer and class 2 doesn't.

class 2 is good, class 1 is bad ## 3 for patient from class 1 The expectation is 5.6864

```
class1_expec = 1+0.9831+0.7609+0.5411+0.9786+0.4227+1
class1_expec
```

```
## [1] 5.6864
```

for patient from class 2 The expectation is 0.8103

```
class2_expec = 0.1165+0.3544+0.2229+0.1165
class2_expec
```

```
## [1] 0.8103
```

4

```
newdf = carcinoma
newdf[,'sum'] = newdf[,1]+newdf[,2]+newdf[,3]+newdf[,4]+newdf[,5]+newdf[,6]+newdf[,7]
newdf
```

```
##      A B C D E F G sum
## 1    1 1 1 1 1 1 1   7
## 2    1 1 1 1 1 1 1   7
## 3    1 1 1 1 1 1 1   7
## 4    1 1 1 1 1 1 1   7
## 5    1 1 1 1 1 1 1   7
## 6    1 1 1 1 1 1 1   7
## 7    1 1 1 1 1 1 1   7
## 8    1 1 1 1 1 1 1   7
## 9    1 1 1 1 1 1 1   7
## 10   1 1 1 1 1 1 1   7
## 11   1 1 1 1 1 1 1   7
## 12   1 1 1 1 1 1 1   7
## 13   1 1 1 1 1 1 1   7
## 14   1 1 1 1 1 1 1   7
## 15   1 1 1 1 1 1 1   7
## 16   1 1 1 1 1 1 1   7
## 17   1 1 1 1 1 1 1   7
## 18   1 1 1 1 1 1 1   7
## 19   1 1 1 1 1 1 1   7
## 20   1 1 1 1 1 1 1   7
## 21   1 1 1 1 1 1 1   7
## 22   1 1 1 1 1 1 1   7
```

```
## 23   1 1 1 1 1 1 1      7
## 24   1 1 1 1 1 1 1      7
## 25   1 1 1 1 1 1 1      7
## 26   1 1 1 1 1 1 1      7
## 27   1 1 1 1 1 1 1      7
## 28   1 1 1 1 1 1 1      7
## 29   1 1 1 1 1 1 1      7
## 30   1 1 1 1 1 1 1      7
## 31   1 1 1 1 1 1 1      7
## 32   1 1 1 1 1 1 1      7
## 33   1 1 1 1 1 1 1      7
## 34   1 1 1 1 1 1 1      7
## 35   1 1 1 1 2 1 1      8
## 36   1 1 1 1 2 1 1      8
## 37   1 2 1 1 1 1 1      8
## 38   1 2 1 1 1 1 1      8
## 39   1 2 1 1 1 1 1      8
## 40   1 2 1 1 1 1 1      8
## 41   1 2 1 1 1 1 1      8
## 42   1 2 1 1 1 1 1      8
## 43   1 2 1 1 1 1 2      9
## 44   1 2 1 1 2 1 1      9
## 45   1 2 1 1 2 1 1      9
## 46   1 2 1 1 2 1 1      9
## 47   1 2 1 1 2 1 1      9
## 48   1 2 1 1 2 1 2     10
## 49   1 2 1 1 2 1 2     10
## 50   1 2 1 1 2 1 2     10
## 51   1 2 1 1 2 1 2     10
## 52   1 2 1 1 2 1 2     10
## 53   2 1 1 1 1 1 1      8
## 54   2 1 1 1 1 1 1      8
## 55   2 1 2 1 2 1 2     11
## 56   2 2 1 1 1 1 1      9
## 57   2 2 1 1 1 1 1      9
## 58   2 2 1 1 1 1 2     10
## 59   2 2 1 1 2 1 1     10
## 60   2 2 1 1 2 1 1     10
## 61   2 2 1 1 2 1 2     11
## 62   2 2 1 1 2 1 2     11
## 63   2 2 1 1 2 1 2     11
## 64   2 2 1 1 2 1 2     11
## 65   2 2 1 1 2 1 2     11
## 66   2 2 1 1 2 1 2     11
## 67   2 2 1 1 2 1 2     11
## 68   2 2 1 1 2 2 2     12
## 69   2 2 1 2 1 1 2     11
## 70   2 2 1 2 2 1 2     12
## 71   2 2 1 2 2 1 2     12
## 72   2 2 1 2 2 2 2     13
## 73   2 2 1 2 2 2 2     13
## 74   2 2 1 2 2 2 2     13
## 75   2 2 2 1 2 1 2     12
## 76   2 2 2 1 2 1 2     12
```

```
## 77   2 2 2 1 2 1 2   12
## 78   2 2 2 1 2 1 2   12
## 79   2 2 2 1 2 1 2   12
## 80   2 2 2 1 2 1 2   12
## 81   2 2 2 1 2 1 2   12
## 82   2 2 2 1 2 1 2   12
## 83   2 2 2 1 2 1 2   12
## 84   2 2 2 1 2 1 2   12
## 85   2 2 2 1 2 1 2   12
## 86   2 2 2 1 2 1 2   12
## 87   2 2 2 1 2 1 2   12
## 88   2 2 2 1 2 2 2   13
## 89   2 2 2 1 2 2 2   13
## 90   2 2 2 1 2 2 2   13
## 91   2 2 2 1 2 2 2   13
## 92   2 2 2 1 2 2 2   13
## 93   2 2 2 2 2 1 2   13
## 94   2 2 2 2 2 1 2   13
## 95   2 2 2 2 2 1 2   13
## 96   2 2 2 2 2 1 2   13
## 97   2 2 2 2 2 1 2   13
## 98   2 2 2 2 2 1 2   13
## 99   2 2 2 2 2 1 2   13
## 100 2 2 2 2 2 1 2   13
## 101 2 2 2 2 2 1 2   13
## 102 2 2 2 2 2 1 2   13
## 103 2 2 2 2 2 2 2   14
## 104 2 2 2 2 2 2 2   14
## 105 2 2 2 2 2 2 2   14
## 106 2 2 2 2 2 2 2   14
## 107 2 2 2 2 2 2 2   14
## 108 2 2 2 2 2 2 2   14
## 109 2 2 2 2 2 2 2   14
## 110 2 2 2 2 2 2 2   14
## 111 2 2 2 2 2 2 2   14
## 112 2 2 2 2 2 2 2   14
## 113 2 2 2 2 2 2 2   14
## 114 2 2 2 2 2 2 2   14
## 115 2 2 2 2 2 2 2   14
## 116 2 2 2 2 2 2 2   14
## 117 2 2 2 2 2 2 2   14
## 118 2 2 2 2 2 2 2   14
```

for class 1(has cancer), $P(sum = 0)$

```
predicted = twoclass$predclass
predicted
```

```
##    [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##   [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1
##   [75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [112] 1 1 1 1 1 1 1
```

```r
count1 = rep(0,8)
for (j in 7:14) {
  for (i in 1:118) {
  if (predicted[i] == 1) {
    if (newdf[i,8] == j) {
      count1[j-6] = count1[j-6]+1
    }
  }
}
}
```

```r
count1
```

```
## [1]  0  0  0  0  9 16 18 16
```

```r
count0 = rep(0,8)
for (j in 7:14) {
  for (i in 1:118) {
  if (predicted[i] == 2) {
    if (newdf[i,8] == j) {
      count0[j-6] = count0[j-6]+1
    }
  }
}
}
count0
```

```
## [1] 34 10  7  8  0  0  0  0
```

```r
result = cbind(count1,count0)
result = result/59
result
```

```
##           count1    count0
## [1,] 0.0000000 0.5762712
## [2,] 0.0000000 0.1694915
## [3,] 0.0000000 0.1186441
## [4,] 0.0000000 0.1355932
## [5,] 0.1525424 0.0000000
## [6,] 0.2711864 0.0000000
## [7,] 0.3050847 0.0000000
## [8,] 0.2711864 0.0000000
```

```r
resultmtx = as.data.frame(result)
#row.names(resultmtx,c(0:7))

resultmtx[,3]=c(0:7)
names(resultmtx)[3] = 'sum'
resultmtx
```

```
##        count1    count0 sum
```

```
## 1 0.0000000 0.5762712    0
## 2 0.0000000 0.1694915    1
## 3 0.0000000 0.1186441    2
## 4 0.0000000 0.1355932    3
## 5 0.1525424 0.0000000    4
## 6 0.2711864 0.0000000    5
## 7 0.3050847 0.0000000    6
## 8 0.2711864 0.0000000    7
```

5

```
predicted = twoclass$predclass
```

```
predicted
```

```
##    [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##   [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [112] 1 1 1 1 1 1 1
```

```
length(predicted)
```

```
## [1] 118
```

```
#relabel of result
for (i in 1:118) {
  if (predicted[i] == 2) {
    predicted[i] = 1
  } else {
    predicted[i] = 2
  }
}
```

```
predicted
```

```
##    [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2
##   [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [112] 2 2 2 2 2 2 2
```

```
table(predicted,carcinoma$A)
```

```
##
## predicted  1  2
##         1 52  7
##         2  0 59
```

```
table(predicted,carcinoma$B)
```

```
##
## predicted  1  2
##         1 38 21
##         2  1 58
```

```
table(predicted,carcinoma$C)
```

```
##
## predicted  1  2
##         1 59  0
##         2 14 45
```

```
table(predicted,carcinoma$D)
```

```
##
## predicted  1  2
##         1 59  0
##         2 27 32
```

```
table(predicted,carcinoma$E)
```

```
##
## predicted  1  2
##         1 46 13
##         2  1 58
```

```
table(predicted,carcinoma$F)
```

```
##
## predicted  1  2
##         1 59  0
##         2 34 25
```

```
table(predicted,carcinoma$G)
```

```
##
## predicted  1  2
##         1 52  7
##         2  0 59
```

F seems to least in agreement with the model since the different predicted number is largest, which is 34