Stat 536 HW6

Dongyang Wang

2022-11-30

Question 1

We first construct all possible models.

```r
snp = c(1167, 1509, 763, 234, 107, 14,
        377, 16, 225, 2, 29, 0,
        186, 179, 130, 19, 11, 2)
snp.array = array(snp,c(2,3,3))
saturated.loglin = loglin(snp.array,margin=list(c(1,2,3)),
                                     fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 0
```

```r
indep.loglin = loglin(snp.array,margin=list(1,2,3),
                                 fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 9.094947e-13
```

```r
X1indepX2X3.loglin = loglin(snp.array ,margin=list(1,c(2,3)),
                                        fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 4.547474e-13
```

```r
X2indepX1X3.loglin = loglin(snp.array ,margin=list(2,c(1,3)),
                                        fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 4.547474e-13
```

```r
X3indepX1X2.loglin = loglin(snp.array ,margin=list(3,c(1,2)),
                                        fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 4.547474e-13
```

```r
X2indepX3givenX1.loglin = loglin(snp.array ,margin=list(c(1,2),c(1,3)),
                                 fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 2.273737e-13
```

```r
X1indepX3givenX2.loglin = loglin(snp.array ,margin=list(c(1,2),c(2,3)),
                                 fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 2.273737e-13
```

```r
X1indepX2givenX3.loglin = loglin(snp.array ,margin=list(c(1,3),c(2,3)),
                     fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 2.273737e-13
```

```
no2nd.loglin = loglin(snp.array ,margin=list(c(1,2),c(1,3),c(2,3)),
                                  fit=TRUE,param=TRUE)
```

```
## 4 iterations: deviation 0.05994974
```

Then we assess their fit.

```
1-pchisq(indep.loglin$lrt,indep.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(X1indepX2X3.loglin$lrt,X1indepX2X3.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(X2indepX1X3.loglin$lrt,X2indepX1X3.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(X3indepX1X2.loglin$lrt,X3indepX1X2.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(X2indepX3givenX1.loglin$lrt,X2indepX3givenX1.loglin$df)
```

```
## [1] 0.552064
```

```
1-pchisq(X1indepX3givenX2.loglin$lrt,X1indepX3givenX2.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(X1indepX2givenX3.loglin$lrt,X1indepX2givenX3.loglin$df)
```

```
## [1] 0
```

```
1-pchisq(no2nd.loglin$lrt,no2nd.loglin$df)
```

```
## [1] 0.4619473
```

Based on above result, we further check model fit for X2indepX3givenX1.loglin and no2nd.loglin to see if u23 is significant.

```
1-pchisq(X2indepX3givenX1.loglin$lrt-no2nd.loglin$lrt,
                X2indepX3givenX1.loglin$df-no2nd.loglin$df)
```

```
## [1] 0.5166232
```

We fail to reject that the term u23 is 0. Thus, we are left with the X2indepX3givenX1.loglin model, shown below.

```
X2indepX3givenX1.loglin
```

```
## $lrt
## [1] 6.857671
##
## $pearson
## [1] 6.354666
##
## $df
## [1] 8
##
## $margin
## $margin[[1]]
```

```
## [1] 1 2
##
## $margin[[2]]
## [1] 1 3
##
##
## $fit
## , , 1
##
##             [,1]      [,2]      [,3]
## [1,] 1176.631 760.3893 99.97963
## [2,] 1515.913 226.8532 14.23392
##
## , , 2
##
##              [,1]        [,2]         [,3]
## [1,] 364.48414 235.545242 30.9706177
## [2,]  15.53013   2.324051  0.1458228
##
## , , 3
##
##             [,1]       [,2]       [,3]
## [1,] 188.8848 122.06544 16.049750
## [2,] 172.5570  25.82278  1.620253
##
##
## $param
## $param$`(Intercept)`
## [1] 3.992946
##
## $param$`1`
## [1]  1.109723 -1.109723
##
## $param$`2`
## [1]  1.5782732  0.4102481 -1.9885213
##
## $param$`3`
## [1]  1.6258730 -1.2505865 -0.3752865
##
## $param$`1.2`
##              [,1]        [,2]        [,3]
## [1,] -0.6109318  0.1205132  0.4904186
## [2,]  0.6109318 -0.1205132 -0.4904186
##
## $param$`1.3`
##              [,1]       [,2]        [,3]
## [1,] -0.6254727  1.079059 -0.4535865
## [2,]  0.6254727 -1.079059  0.4535865
```

Therefore, from the analysis, we learned that SNP1 is independent of SNP2 given disease. But SNP1 and SNP2 are related to each other.

The interpretations of individual models are as follows:

First, we specify the X1 as disease, X2 as SNP1, X3 as SNP2. The interpretation of the models is as follows.

First model is saturated model and we consider all possible higher order interactions: that is all three variables can interact each other.

Second model is complete independence model. That is, all variables act on their own.

Third model: This is the model of independence of X1 and {X2, X3}. Meaning: disease independent of SNP1, SNP2.

Fourth model: This is the model of independence of X2 and {X1, X3}. Meaning: SNP1 independent of disease, SNP2.

Fifth model: This is the model of independence of X3 and {X1, X2}. Meaning: SNP2. independent of SNP1, disease.

Sixth model: This is the model of conditional independence of X2 and X3 given X1. Meaning: SNP1 independent of SNP2 given disease.

Seventh model: This is the model of conditional independence of X1 and X3 given X2. Meaning: disease. independent of SNP2 given SNP1.

Eighth model: This is the model of conditional independence of X1 and X2 given X3. Meaning: SNP1 independent of disease given SNP2.

Last model: this is the no second order interaction model.