Stat 536 HW5

Dongyang Wang

2022-11-10

Question 1

```
df1 = c(2037,1757,958,218)
df1 = array(df1, c(2,2))

indep_mod1 <- loglin(df1,margin=list(1,2),fit=TRUE,param=TRUE)
```

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

The independence model does not seem to fit well since the prediction is a bit off. We fit the saturated model instead.

```
sat_mod1 <- loglin(df1,margin=list(c(1,2)),fit=TRUE,param=TRUE)
```

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

Question 2

$\log \frac{P(X_1=Yes|X_2=k)}{P(X_1=No|X_2=k)} = u_{1(1)} - u_{1(2)} = 0.2081879 + 0.2081879 = 0.4163758$. The interpretation is that the odds of yes over no are $\exp(0.4163758) = 1.516456$.

```
exp(0.4163758)
```

```
## [1] 1.516456
```

Question 3

Under the independence model, we reject the null.and there is indeed a difference. The alternative hypotheses include: BB is different from Bb only; BB is different from bb only; BB is different from Bb and bb.

```
1- pchisq(indep_mod1$lrt, indep_mod1$df)
```

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

Question 1

The asymptotic result is calculated by subtracting the lrt's and df's.

```
df2 = c(2037,1757,631, 18, 327, 200)
df2 = array(df2, c(2,3))

indep_mod2 <- loglin(df2,margin=list(1,2),fit=TRUE,param=TRUE)
```

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

```
sat_mod2 <- loglin(df2,margin=list(c(1,2)),fit=TRUE,param=TRUE)
```

```
## 2 iterations: deviation 2.842171e-14
```

```
1-pchisq(indep_mod2$lrt-sat_mod2$lrt,indep_mod2$df-sat_mod2$df)
```

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

Question 2

```
1-pchisq(indep_mod2$lrt,indep_mod2$df)
```

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

Question 3

$\log \frac{P(X_1=Yes|X_2=k)}{P(X_1=No|X_2=k)} = u_{1(1)} - u_{1(2)} = 0.2081879 + 0.2081879 = 0.4163758$. The interpretation is that the odds of yes over no are $\exp(0.4163758) = 1.516456$.