Stat 536 HW4

Dongyang Wang

2022-10-29

Question 1

```
rm(list = ls())
table1 <- c(28, 656, 28+656, 18, 658, 18+658, 28+18, 656+658, 28+18+656+658)
matrix_d <- matrix(table1,byrow = TRUE, nrow = 3)
colnames(matrix_d) <- c("Yes", "No", "Row Total")
rownames(matrix_d) <- c("Placebo", "Aspirin", "Col Total")
matrix_d
```

```
##           Yes   No Row Total
## Placebo    28  656       684
## Aspirin    18  658       676
## Col Total  46 1314      1360
```

Grand total is 1360.

Question 2

By Frechet lower and upper bounds, (1,1) has lower bound 0 and upper bound 46. (1,2) has lower bound 638 and upper bound 684. (2,1) has lower bound 0 and upper bound 46. (2,2) has lower bound 630 and upper bound 676.

Question 3

Since the smallest range for the cell entries above involves (1,1), the number of T is equal to the number of possible values of (1,1), given current row and column total constraints. In our case, it is 47 because (1,1) can take on any integer between 0 and 46.

The notation with x being the count in (1,1). T = {(x, 684-x, 46 - x, 630+x), $x \in \{0,1,\ldots 46\}$}.

Question 4

The following is saturated log linear fit.

```
df = c(28,18,656,658)
df = array(df, c(2,2))
loglin(df,margin=list(c(1,2)),fit=TRUE,param=TRUE)$fit
```

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

```
##      [,1] [,2]
## [1,]   28  656
## [2,]   18  658
```

Question 5

The following is the log linear fit of independence.

```
loglin(df,margin=list(1,2),fit=TRUE,param=TRUE)$fit
```

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

```
##           [,1]     [,2]
## [1,] 23.13529 660.8647
## [2,] 22.86471 653.1353
```

Question 6

```
x2 <- 0
for (i in 1:2){
  for (j in 1:2){
    core <- matrix_d[i,3] * matrix_d[3,j]/matrix_d[3,3]
    x2 <- x2 + ((matrix_d[i,j]-core)/sqrt(core))^2
  }
}
pval_6 <- 1 - pchisq(x2, df=1)
pval_6
```

```
## [1] 0.1444434
```

Fail to reject the null that aspirin and myocardial infarction are independent.

Question 7

```
d <- 0
for (i in 1:2){
  for (j in 1:2){
    core <- matrix_d[i,3] * matrix_d[3,j]/matrix_d[3,3]
    d <- d + 2*matrix_d[i,j] * log(matrix_d[i,j]/core)
  }
}
pval_7 <- 1 - pchisq(d, df=1)
pval_7
```

```
## [1] 0.1428159
```

Fail to reject the null that aspirin and myocardial infarction are independent.

Question 8

X1: Aspirin Use

X2: Myocardial Infarction

The null: X1 and X2 are independent

The alternative: X1 and X2 are not independent

```
fisher.test(df, simulate.p.value = TRUE, B = 1e5)
```

```
##
##  Fisher's Exact Test for Count Data
##
```

```
## data:  df
## p-value = 0.1768
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8236804 3.0256940
## sample estimates:
## odds ratio
##   1.559785
```

Question 9

With the model assuming independence, we have

```
loglin(df,margin=list(c(1,2)),fit=TRUE,param=TRUE)$param$`1`
```

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

```
## [1]  0.1096972 -0.1096972
```

where u1(2) - u1(1) = -0.1096972 -0.1096972 = -0.2193944. This is conditional distribution.

Question 10

Based on the results from 6,7,8, we know that we fail to reject the null that aspirin and myocardial infarction are independent. That means, the effect of aspirin on the occurrence of myocardial infarction is not significant.