Math 490 HW #16

Maxwell Levin

March 30, 2018

Question 1.

Use our class data to carry out the chi-square test on whether gender and sleep are independent. Follow the 4-step format.

Step 1: State the null and alternative hypotheses

Since we are asked to test whether gender and sleep are independent attributes, a good place to start for our null hypothesis is that gender and sleep are not independent. This means that our alternative hypothesis is that gender and sleep are independent attributes.

Step 2: Summarize the data using a contingency table

This problem is simple enough that we do not necessarily need R to generate our contingency table, but nonetheless I have written the following code in R as good practice and to save time in the future should I need to create a larger contingency table later on.

```r
# Import our class data
our_data = read.table("math490.R", header=TRUE)
attach(our_data)

# Splice the data for our contingency table
early_m = our_data[ which(sleep_type == 'early' & gender == 'm'), ]
early_f = our_data[ which(sleep_type == 'early' & gender == 'f'), ]
night_m = our_data[ which(sleep_type == 'night' & gender == 'm'), ]
night_f = our_data[ which(sleep_type == 'night' & gender == 'f'), ]

# Calculate the totals (num_early + num_night and num_m + num_f should both be 22)
num_early = nrow(early_m) + nrow(early_f)
num_night = nrow(night_m) + nrow(night_f)
num_m = nrow(our_data[ which(gender == 'm'), ])
num_f = nrow(our_data[ which(gender == 'f'), ])

obs = c(nrow(early_m), nrow(early_f), nrow(night_m), nrow(night_f)) # I'll use this later

# Create the table
makeTable = function(value, totals) {
  # Must have exactly 1 degree of freedom
  # Value is the top left value in the table
  # Totals are the sum of row elements in decending order, then the columns in ascending order.
  v2 = totals[1] - value
  v3 = totals[3] - value
  v4 = totals[2] - v3
  ans = matrix(c(value, v2, totals[1],
                 v3, v4, totals[2], totals[3],
                 totals[4], totals[1] + totals[2]),
               ncol=3)
```

```r
  colnames(ans) = c("Male", "Female", "Total")
  rownames(ans) = c("Early", "Night", "Total")
  ans
}
totals = c(num_early, num_night, num_m, num_f)
as.table(makeTable(5, totals))
```

```
      Male Female Total
Early    5      5    10
Night    5      7    12
Total   10     12    22
```

Now we run the following code in R to compute the expected frequencies

```r
total = num_m + num_f
row1 = c(num_early * num_m / total, num_early * num_f / total)
row2 = c(num_night * num_m / total, num_night * num_f / total)
exp = c(row1, row2) # I'll use this later
exp_table = matrix(c(row1, row2), ncol=2)
colnames(exp_table) = c("Male", "Female")
rownames(exp_table) = c("Early", "Night")
as.table(exp_table)
```

```
          Male   Female
Early 4.545455 5.454545
Night 5.454545 6.545455
```

Step 3: Compute the chi-square statistic

We know from class that we can compute this by

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(Obs_{i,j} - Exp_{i,j})^2}{Exp_{i,j}}.$$

We run the following code in R to compute our chi-square statistic with 1 degree of freedom

```r
chi_stat = sum(((obs - exp)^2) / exp)
chi_stat
```

```
[1] 0.1527778
```

Step 4: Assess the P-value from our chi-square statistic

We can compute our P-value by running the following code in R

```r
1 - pchisq(chi_stat, 1)
```

```
[1] 0.6958948
```

We see that this is much bigger than a 5% significance level. Thus we do not have enough evidence to reject the null hypothesis that sleep and gender are not independent.

Question 2.

Use our class data to carry out Fisher's exact test on whether gender and sleep are independent. List all possible contingency tables and compute the P-value.

To do this we must list all the possible contingency tables. We can do this in R quite easily:

```
makeTable = function(value, totals) {
  # Must have exactly 1 degree of freedom
  # Value is the top left value in the table
  # Totals are the sum of row elements in decending order, then the columns in ascending order.
  v2 = totals[1] - value
  v3 = totals[3] - value
  v4 = totals[2] - v3
  ans = matrix(c(value, v2, v3, v4), ncol=2)
  colnames(ans) = c("Male", "Female")
  rownames(ans) = c("Early", "Night")
  ans
}

fish = numeric(11)
for (i in 1:11) {
  matAns = makeTable(i-1, totals)
  fish[i] = choose(10, i-1) * choose(12, matAns[4]) / choose(22, 10)
  print(as.table(matAns))
  writeLines('This table has a probibility of:')
  print(fish[[i]])
  writeLines('')
}
```

```
      Male Female
Early    0     10
Night   10      2
This table has a probibility of:
[1] 0.0001020651

      Male Female
Early    1      9
Night    9      3
This table has a probibility of:
[1] 0.003402171

      Male Female
Early    2      8
Night    8      4
This table has a probibility of:
[1] 0.03444698

      Male Female
Early    3      7
Night    7      5
This table has a probibility of:
[1] 0.1469738

      Male Female
```

```
Early    4      6
Night    6      6
This table has a probibility of:
[1] 0.3000714

      Male Female
Early    5      5
Night    5      7
This table has a probibility of:
[1] 0.3086449

      Male Female
Early    6      4
Night    4      8
This table has a probibility of:
[1] 0.1607526

      Male Female
Early    7      3
Night    3      9
This table has a probibility of:
[1] 0.04082605

      Male Female
Early    8      2
Night    2     10
This table has a probibility of:
[1] 0.00459293

      Male Female
Early    9      1
Night    1     11
This table has a probibility of:
[1] 0.0001855729

      Male Female
Early   10      0
Night    0     12
This table has a probibility of:
[1] 1.546441e-06
```

We now must compute the sum of all the P-values (probabilities) less than or equal to the one we observed. We do this with the following command:

```
sum(fish[fish <= 0.3086449])
```

```
[1] 0.6913551
```

Thus we see that our P-value is pretty close to what we calculated in the first question. Thus both the chi-square test and Fisher's exact test tell us that we do not have sufficient evidence to reject the null hypothesis.