PS 12

Aditya Sanjay Mhaske

2023-04-21

Question 1

Question 1.a.)

Null hyothesis = Ho : obsereverd proportion = expected proportions Alternate hypothesis = H1 : obsereverd proportion = expected proportions

```
obs = c(121, 84, 118, 226, 226, 123)
n = sum(obs)
exp = n * c(0.13, 0.14, 0.13, 0.24, 0.20, 0.16)
df = (6-1) - 0
g2 = 2*sum(obs * log(obs/exp))
g_pval = 1 - pchisq(g2,df)
g_pval
```

```
## [1] 1.141029e-05
```

Since the p value is less than significance level, we can reject the null hypothesis

Question 1.b)

```
library(infer)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.2.1     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
x2 = sum((obs - exp)^2/exp)
x_pval = 1 - pchisq(x2,df)
vec = rep(as.character(1:6), obs)
df = data.frame(vec)
null_dist <- df |>
  specify(response = vec) |>
  hypothesize(null = "point", p = c("1" = 0.13,
                                    "2" = 0.14,
```

```
                                    "3" = 0.13,
                                    "4" = 0.24,
                                    "5" = 0.20,
                                    "6" = 0.16)) |>
  generate(reps = 2500, type = "draw") |>
  calculate(stat = "Chisq")

null_dist |>
  get_p_value(obs_stat = x2, direction = "greater")
```

```
## Warning: Please be cautious in reporting a p-value of 0. This result is an
## approximation based on the number of `reps` chosen in the `generate()` step. See
## `?get_p_value()` for more information.
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1       0
```

Since the p value is less than significance level, we can reject the null hypothesis

Question 1.c) From the expected

```
obs
```

```
## [1] 121  84 118 226 226 123
```

```
exp
```

```
## [1] 116.74 125.72 116.74 215.52 179.60 143.68
```

From the expected, we can see that the the observed yellow candies are much less than expected and the observed orange candies are much more than expected. So we increase the proportion of orange candies and decrease the expected proportion of yellow candies.

```
exp = n * c(0.13, 0.11, 0.13, 0.24, 0.23, 0.16)
df = (6-1) - 0
g2 = 2*sum(obs * log(obs/exp))
g_pval = 1 - pchisq(g2,df)
g_pval
```

```
## [1] 0.1613472
```

Now we fail to reject the null, and hence change the conclusion of previous test

Question 2

Question 2.a.)

```
n = 16
p = 0.29
exp = c(sum(dbinom(0:1,n,p)), dbinom(2:8,n,p), sum(dbinom(9:16,n,p))) * 1000
exp
```

```
## [1]  31.42165  83.48225 159.12578 211.23387 207.06870 155.05849  90.47678
## [8]  41.57472  20.55776
```

```
obs = c(30, 93, 159, 184, 195, 171, 92, 45, 31)
obs
```

```
## [1]  30  93 159 184 195 171  92  45  31
```

```
df = (9-1)-0
g2 = 2*sum(obs * log(obs/exp))
g_pval = 1 - pchisq(g2,df)
g_pval
```

```
## [1] 0.1525781
```

We fail to reject the null, i.e., the algorithm is not working as intended.

Question 2.b)

```
x2 = sum((obs - exp)^2/exp)
x_pval = 1 - pchisq(x2,df)
x_pval
```

```
## [1] 0.1257997
```

```
color.vec = rep(as.character(1:9),obs)
df = data.frame(color.vec)

null_dist <- df |>
  specify(response = color.vec) |>
  hypothesize(null = "point", p = c("1" = sum(dbinom(0:1,n,p)),
                                    "2" = dbinom(2,n,p),
                                    "3" = dbinom(3,n,p),
                                    "4" = dbinom(4,n,p),
                                    "5" = dbinom(5,n,p),
                                    "6" = dbinom(6,n,p),
                                    "7" = dbinom(7,n,p),
                                    "8" = dbinom(8,n,p),
                                    "9" = sum(dbinom(9:16,n,p)))) |>
  generate(reps = 2500, type = "draw") |>
  calculate(stat = "Chisq")
null_dist |>
  get_p_value(obs_stat = x2, direction = "greater")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.125
```

We fail to reject the null, i.e., the algorithm is not working as intended.

Question 2.c)

```r
mu = sum(obs*c(1:9)) / 16000
mu
```

```
## [1] 0.2945625
```

No, there is no change in degree of freedom as there is no change on the constraints on variables.

Question 3

Question 3.)

Null hyothesis = histological type and response to treatment are independent Alternate hypothesis = histological type and response to treatment are dependent

Question 3.a)

```r
obs = matrix(c(74, 18, 12, 68, 16, 12, 154, 54, 58, 18, 10, 44), nrow = 4, ncol = 3)
obs
```

```
##      [,1] [,2] [,3]
## [1,]   74   16   58
## [2,]   18   12   18
## [3,]   12  154   10
## [4,]   68   54   44
```

```r
exp = rowSums(obs)%o%colSums(obs)/sum(obs)
exp
```

```
##          [,1]     [,2]     [,3]
## [1,] 47.31599 64.92193 35.76208
## [2,] 15.34572 21.05576 11.59851
## [3,] 56.26766 77.20446 42.52788
## [4,] 53.07063 72.81784 40.11152
```

```r
df = (4 - 1)*(3 - 1)
g2 = sum(2*obs*log(obs/exp))
g2
```

```
## [1] 241.733
```

```r
1 - pchisq(g2, df)
```

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

We can reject the null. Hence, histological type and response to treatment are dependent

Question 3.b)

```r
x2 = sum((obs - exp)^2/exp)
rownames(obs) <- c("LP", "NS", "MC", "LD")
colnames(obs) <- c("Positive", "Partial", "None")
obs
```

```
##    Positive Partial None
## LP       74      16   58
## NS       18      12   18
## MC       12     154   10
## LD       68      54   44
```

```
df.obs = as.data.frame(obs)
data2 <- df.obs |>
  rownames_to_column("histologicalType") |>
  pivot_longer(cols=c('Positive', 'Partial', 'None'),
               names_to='response',
               values_to='count') |>
  rowwise() |>
  mutate(count = list(1:count)) |>
  unnest(count) |>
  select(-count)



null_dist <- data2 |>
  specify(response ~ histologicalType) |>
  hypothesize(null = "independence") |>
  generate(reps = 2500, type = "permute") |>
  calculate(stat = "Chisq")

null_dist |>
  get_p_value(obs_stat = x2, direction = "greater")
```

```
## Warning: Please be cautious in reporting a p-value of 0. This result is an
## approximation based on the number of ‘reps‘ chosen in the ‘generate()‘ step. See
## ‘?get_p_value()‘ for more information.
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1       0
```

We can reject the null. Hence, histological type and response to treatment are dependent