Stat 120C Homework 6

Hyundo Im

June 6 2017

Problem 1

let's suppose: p_1 = proportion of Bb or bb in the population of diabetic adults p_2 = Proportion of Bb or bb in the population of Normal adults

	Diabetic	Normal
Bb or bb	3	1
BB	5	6

In Fisher's Exact Test, we do the hypothesis test that: $H_0 : p_1 = p_2$ vs $H_a : p_1 \neq p_2$ To get test statistic, we use hypergeometric distribution.

Recall we can write: $Pr(N_{11} = n_{11})$ $\frac{\binom{4}{n_{11}}\binom{11}{8-n_{11}}}{\binom{15}{8}}$

Where $n_{11} = 0, 1, 2, 3, 4$

```
N11<-cbind(N11=0:4,prob=dhyper(0:4,4,11,8))
N11
```

```
##      N11       prob
## [1,]   0 0.02564103
## [2,]   1 0.20512821
## [3,]   2 0.43076923
## [4,]   3 0.28717949
## [5,]   4 0.05128205
```

p-value is the sum of every extreme probability. To get that sum, I use R code below.

```
fisher.test(matrix(c(3,5,1,6),2,2))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  matrix(c(3, 5, 1, 6), 2, 2)
## p-value = 0.5692
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.1907926 219.5351657
## sample estimates:
## odds ratio
##   3.309695
```

Since our p-value is .5692>.05, or it is too large, we fail to reject the Null hypothesis. We do not have enough evidence to claim that there is the proportion of Bb or bb in the population of diabetic adults is different from the proportion of Bb or bb in the population of Normal adults.

Problem 2

(a)

$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{4} \frac{(Obs_{ij} - Exp_{ij})}{Exp_{ij}}$

First, we need a table of Expected counts Expected counts $= \pi_{ij} = \frac{n_{i.} n_{.j}}{n}$

Expected counts table is:

	O	A	B	AB	
Rh+	27.26	94.35	49.61	22.77	244
Rh-	17.73	21.65	11.39	5.23	56
	95	116	61	28	300

Then, Chi-sq test statistic would be 8.6

(b)

Confirm with R code:

```
rhp = c(82,89,54,19)
rhn = c(13,27,7,9)
chi_<-chisq.test(rbind(rhp,rhn), correct= FALSE)
chi_
```

```
##
##  Pearson's Chi-squared test
##
## data:  rbind(rhp, rhn)
## X-squared = 8.6037, df = 3, p-value = 0.03505
```

Since our test statistic is large enough and p-value is 0.035<0.05, we can reject the null hypothesis at 5% significant level. We have enough evidence to claim two classifications of blood type are not independent.

(c)

To calculate $2(log(L_1 - log(L_0)))$, we need to find out L_1 and L_0.

$L_1 = \frac{n!}{n_{11}! n_{12}! \ldots n_{24}!} \prod_{i=1}^{2} \prod_{j=1}^{4} p_{ij}^{n_{ij}}$

$l_1 = log(L_1) = \sum_i^2 \sum_j^4 n_{ij} log(p_{ij}) + C$

$L_0 = \frac{n!}{n_{11}! n_{12}! \ldots n_{24}!} \prod_{i=1}^{2} \prod_{j=1}^{4} [p_{i.} p_{.j}]^{n_{ij}}$

$l_0 = log(L_0) = \sum_i^2 \sum_j^4 n_{ij} [log(p_{i.}) + log(p_{.j})]$

Then, $2(log(L_1) - log(L_0)) = 2(l_1 - l_0)$

$= 2(\sum_i \sum_j n_{ij} [log(p_{ij}) - log(p_{i.}) - log(p_{j.})]), where p_{ij} = \frac{n_{ij}}{n}, p_{i.} = \frac{n_{i.}}{n}, p_{.j} = \frac{n_{.j}}{n}$

$= 2(\sum_i \sum_j n_{ij} (log(\frac{n_{ij}/n}{n_{i.} n_{.j}/n^2})))$

$= 2(\sum_i \sum_j n_{ij} (log(\frac{n_{ij}}{n_{i.} n_{.j}/n})))$

$= 2(\sum_i \sum_j n_{ij}(log(\frac{Obj_{ij}}{Exp_{ij}})))$

To shorten the calculations, I used the R code below

```
exp_<-chi_$expected
rh<-rbind(rhp,rhn)
val<-rh/exp_
sum(2*(rh*log(val)))
```

```
## [1] 8.447251
```

Therefore, $2(\sum_i \sum_j n_{ij}(log(\frac{Obj_{ij}}{Exp_{ij}}))) = 8.45$

(d)

Comparing to Pearson's chi-square statistic, test statistic in (c) is smaller (8.45<8.6).

```
pchisq(8.45,df=3,lower.tail = FALSE)
```

```
## [1] 0.03757187
```

Our p-value is 0.038<0.05, so we reject the null hypothesis at 5% significant level. We have enough evidence to claim that two classifications of blood type are not independent.

Problem 3

Under $H_0 : p_{i1} = p_{i2} = ... = p_{iJ} = p_i$ for i = 1,. . . ,I Under H_0, with constraint $\sum_{i=1}^{I} p_{ij} = 1$,

$L_0(p_1, ..., p_I) = \prod_{j=1}^{J} \binom{n_j}{n_{1j}, n_{2j}, ..., n_{Ij}} p_{1j}^{n_{1j}} ... p_{Ij}^{n_{Ij}}$

$L_0(p_1, ..., p_I) \propto p_1^{n_{1.}} \cdot p_2^{n_{2.}} \cdot ... \cdot (1 - \sum_{i=1}^{I-1} p_i)^{n_{I.}}$ because $p_I = 1 - \sum_{i=1}^{I-1} p_i$

Then, loglikelihood

$l_0 = C + n_{1.}log(p_1) + n_{2.}log(p_2) + ... + n_{I.}log(1 - \sum_{i=1}^{I-1} p_i)$

Therefore, we can express as: $\frac{\partial l_0}{\partial p_1} = \frac{n_{1.}}{p_1} - \frac{n_{I.}}{1-\sum_{i=1}^{I-1} p_i} = \frac{n1.}{p_1} - \frac{n_{I.}}{p_I} = 0$

Therefore, $\hat{p_1} = \frac{p_I n_{1.}}{n_{I.}}$.

It implies that: $\frac{\hat{p_i}}{\hat{p_j}} = \frac{n_{i.}}{n_{.j}}$

We can rewrite as: $\hat{p_j} = \frac{n_{j.}\hat{p_i}}{n_{i.}} \rightarrow \sum_{j=1}^{J} \hat{p_j} = \frac{n_{..}\hat{p_i}}{n_{i.}} = 1$

Therefore, $\hat{p}_{iMLE} = \frac{n_{i.}}{n}$

Problem 4

It is a matched-pairs design. The given table is correlated table of severe cold at age 12 and severe cold at age 10. We can conduct the test with null hypothesis of probabilities of severe cold are the same among age of 10 and 12.

Another word, $H_0 : \pi_{12} = \pi_{21}$.

Recall that under the null hypothesis, mle's f the cell probabilities are: $\hat{\pi_{11}} = \frac{n_{11}}{n}, \hat{\pi_{21}} = \hat{\pi_{12}} = \frac{n_{12}+n_{21}}{2n}, \hat{\pi_{22}} = \frac{n_{22}}{n}$

Recall from the class, the formula for chi-sq in matched pair is: $\chi^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}}$

	Severe colds at age 12 Yes	Severe colds at age 12 No
Severe Colds at age 10 Yes	200	165
Severe Colds at age 10 No	235	600

$\chi^2 = \frac{(165-235)^2}{165+235} = 12.25$

Full model's parameters: $\pi_{11}, \pi_{12}, \pi_{21}$ Reduced model's parameters: π_{11}, π_{12}a

Therefore, degrees of freedom is 3-2=1 p-value of $\chi^2 = 12.25$ at df=1 is:

```
pchisq(12.25,1,lower.tail = FALSE)
```

```
## [1] 0.0004652582
```

Since our p-value is approximately zero, we can reject the null hypothesis that we can conclude we have enough evidence to claim the probabilities of having colds at age 10 and 12 are not the same.