Stat 261 - Assign # 4 - Spring 2018

Calculations were done in R - see attached R script

$\overline{20} + \frac{\text{Presentation}}{2} = \frac{\quad}{22}.$

1) 5 Let (f_1, f_2, f_3) be the frequencies of the 3 categories (MM, FF, MF or FM)

(i) Then the Basic model is $(f_1, f_2, f_3) \sim$ Multinomial $(n = 28, p_1, p_2, p_3)$.

There are $k = 3-1$ functionally indep. unknown parameters.

We derived the MLE's in class: $\hat{p}_i = \frac{f_i}{n}$

(ii) The hypothesis is: $H_0: p_1 = \frac{3}{14}, \; p_2 = \frac{3}{14}, \; p_3 = \frac{8}{14}$.

There are $q = 0$ unknown parameters under H_0.

(iii) LRS $D = 2\left\{ l(\hat{p}_1, \hat{p}_2, \hat{p}_3) - l\left(\frac{3}{14}, \frac{3}{14}, \frac{8}{14}\right) \right\}$.

where $l(p_1, p_2, p_3) = \sum_{i=1}^{3} f_i \ln p_i$

We showed in class that $D = 2 \sum_{i=1}^{3} f_i \ln\left(\frac{f_i}{e_i}\right)$, where e_i is the expected frequency under H_0.

Here $d_{obs} = 2\left\{ 9 \ln \frac{9}{28 \times \frac{3}{14}} + 4 \ln \frac{4}{28 \times \frac{3}{14}} + 15 \ln \frac{15}{28 \times 8/14} \right\}$

$\boxed{= 2.12.}$

p-value $\simeq P\left(\chi^2_{(k-q=2)} \geq 2.12\right) = \boxed{.35}$

There is no evidence against the null hypothesis. (pvalue = .35)

P [The data are consistent with the hypothesis that $p_1 = \frac{3}{14}$, $p_2 = 3/14$ and $p_3 = 8/14$. $(p = .35)$]

2 (a) 5

(i) Let Y_i = # germinate from colour $i = 1,2,3,4$ (red, white, blue, yellow).

The Basic model is $Y_i \sim \text{Binomial}(n_i = 100, p_i)$

$p_i = P(\text{germination for colour } i)$

There are $k=4$ functionally indep. unknown parameters; $k=4$.

We derived the MLE's in class: $\hat{p}_i = y_i/n$.

(ii) The hypothesis is: $H_0: p_1 = p_2 = p_3 = p_4 = .80$.

There are $q=0$ unknown parameters under H_0.

(iii) LRS, $D = 2\{\ell(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) - \ell(.80, .80, .80, .80)\}$.

$\ell(p_1, p_2, p_3, p_4) = \sum_{i=1}^{4} [y_i \ln p_i + (n_i - y_i)\ln(1-p_i)]$.

$\boxed{d_{obs} = 13.81}$ pvalue $= P\{\chi^2_{(4)} \geq 13.81\} = \boxed{.008}$

There is very strong evidence against the null hypothesis.

P [The data are not consistent with the hypothesis that the germination rate is 80% for all colours. (p = .008)]

(b) 5 (i) The Basic model is the same as part (a)

(ii) The hypothesis is that: $H_0: p_1 = p_2 = p_3 = p_4 = p$

There is $q=1$ parameter to estimate.

$$\ell_H(p) = \ell(p,p,p,p) = \sum_{i=1}^{4} [y_i \ln p + (n_i - y_i)\ln(1-p)].$$

$$= \left(\sum_1^4 y_i\right)\ln p + \left[\sum_1^4 (n_i - y_i)\right]\ln(1-p)$$

$$\frac{d\ell_H}{dp} = \frac{\sum_1^4 y_i}{p} - \frac{\sum_1^4 (n_i - y_i)}{1-p} \Rightarrow \boxed{\tilde{p} = \frac{\sum_1^4 y_i}{\sum_1^4 n_i} = \frac{297}{400} = .7425}$$

(iii) LRS: $D = 2\{ \ell(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) - \ell(.7425, .7425, .7425, .7425) \}$.

$d_{obs} = \boxed{6.06}$ pvalue $= P\{ \chi^2_{(k-q=3)} \geq 6.06 \} = \boxed{.11}$

$\boxed{\text{There is no evidence against the null hypothesis.}}$

(P) [The data are consistent with the hypothesis of a common germination rate for all colours, estimated as 74.25%. ($p = .11$)]

3 (5). (i). Let f_{ij} = frequency of category (i, j).
The Basic model is $(f_{11}, f_{12}, f_{21}, f_{22}) \sim$ Multinomial $(n = 604, p_{11}, p_{12}, p_{21}, p_{22})$.

There are $k = 4 - 1 = 3$ functionally indep. unknown parameters.
We derived the MLE's in class: $\hat{p}_{ij} = \dfrac{f_{ij}}{n}$

(ii) H_0: $p_{ij} = \alpha_i \beta_j$ where $\alpha_i = P\{ i^{th} \text{ smoking status} \}$
$\beta_j = P\{ j^{th} \text{ disease status} \}$.
hypothesis of independence. $q = 2$ parameters to estimate
We derived the MLE's in class: $\tilde{\alpha}_i = \dfrac{r_i}{n}$ $\beta_j = \dfrac{c_j}{n}$.

where $r_i = f_{i1} + f_{i2}$ the row totals.
$c_j = f_{1j} + f_{2j}$ the column totals

and $\tilde{p}_{ij} = \tilde{\alpha}_i \tilde{\beta}_j = \dfrac{r_i c_j}{n^2}$

$\tilde{e}_{ij} = n \tilde{p}_{ij} = \dfrac{r_i c_j}{n}$

(iii) LRS : $D = 2\{ l(\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{21}, \hat{p}_{22}) - l(\tilde{p}_{11}, \tilde{p}_{12}, \tilde{p}_{21}, \tilde{p}_{22}) \}$

$$D = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} \left(f_{ij} \ln \frac{f_{ij}}{e_{ij}} \right).$$

$d_{obs} = \boxed{6.34}$ $\quad$ pvalue $= P(\chi^2_{(1)} \geq 6.34) = \boxed{.012}$

There is evidence against the hypothesis of independence.

P The data provides evidence against the hypothesis that disease classification is independent of smoking status. $p = .012$. The estimated probabilities of cancer are:

$72/469 = .15$ for smokers. and

$10/135 = .07$ for nonsmokers.

Aside: The estimated frequencies under H_0 are in the table (e_{ij}).

	Cancer	Other	Total
Smoker	72 (63.67)	397 (405.33)	$r_1 = 469$
Non-smoker	10 (18.33)	125 (116.67)	$r_2 = 135$
Total	$c_1 = 82$	$c_2 = 522$	604.

Note: The computations for the LRS here are very sensitive to rounding. For accurate results, keep all digits and compute using R.

Assignment 4 - R code

Mary Lesperance

March 12, 2018

Solutions for Lecture Assignment #4, March 2018

1. Multinomial simple hypothesis

```
freq <- c(9, 4, 15)
sum(freq)

## [1] 28

ell <- function(p,freq){
  # Multinomial log-likelihood
  # freq = frequencies;  p = probabilities
  sum(freq*log(p))
}

LRS <- function(p0, phat,freq){
  #Likelihood ratio statisitic
  # freq=frequencies; p0 = H0 probabilities
  # phat = MLE
  2*(ell(phat,freq) - ell(p0,freq))
}

dobs <- LRS(c(3,3,8)/14,freq/sum(freq),freq)   #LRS observed
dobs

## [1] 2.118495

1-pchisq(dobs,2)  #pvalue

## [1] 0.3467165
```

2. Four binomials

(a) H0: p_i = .80, i=1,2,3,4

```
y<-c(76, 66, 81, 74)
n<-rep(100,4)

ell<-function(y,n,p){
  # Binomial log-likelihood
  # y = number of successes; n=size; p=probabilities
  sum(y*log(p) + (n-y)*log(1-p))
}
```

```
LRS<-function(phat,p0,y,n){
  #Likelihood ratio statistic, phat=MLE, p0=H0 values
  # y = number of successes; n=size
  2*(ell(y,n,phat)-ell(y,n,p0))
}

phat<-y/n  #MLE under BASIC model

p0 <- rep(.8, 4)   # H0 under (a)
D<-LRS(phat,p0,y,n)    #observed LRS
D

## [1] 13.81242

1-pchisq(D,4)    #p-value

## [1] 0.00791843
```

(b) H0: p_1 = p_2 = p_3 = p_4 = p

```
p0<-sum(y)/sum(n)  #MLE under (b) H0
p0

## [1] 0.7425

D<-LRS(phat,p0,y,n)    #observed LRS
D

## [1] 6.061518

1-pchisq(D,4-1)    #p-value

## [1] 0.1086553
```

3. Test of Independence

```
freq <- matrix(c(72, 397, 10, 125), nrow=2, byrow=TRUE)
freq

##      [,1] [,2]
## [1,]   72  397
## [2,]   10  125

sum(freq)

## [1] 604

ell <- function(p,freq){
  # Multinomial log-likelihood
  # freq = frequencies;  p = probabilities
  sum(freq*log(p))
}

LRS <- function(p, phat,freq){
```

```
  #Likelihood ratio statistic, phat=MLE, p=H0 values
  # freq=frequencies
  2*(ell(phat,freq) - ell(p,freq))
}

rsum <- rowSums(freq)
csum <- colSums(freq)
rsum

## [1] 469 135

csum

## [1]  82 522

eij <- outer(rsum, csum)/sum(freq)  # expected freq - see help for outer
eij                                 # eij= r_i * c_j/n

##          [,1]     [,2]
## [1,] 63.67219 405.3278
## [2,] 18.32781 116.6722

# estimated probs under H0 are eij/sum(freq)

dobs <- LRS(c(eij)/sum(freq),c(freq)/sum(freq),c(freq))    #LRS observed
dobs

## [1] 6.33659

1-pchisq(dobs,1)  #pvalue, df=(#rows - 1)(#cols - 1)

## [1] 0.0118272
```

Optional Pearson's chi-square test

```
GOF <- sum((freq-eij)^2/eij)
GOF

## [1] 5.638738

1-pchisq(GOF,1)

## [1] 0.01756784
```