Homework Set #9 Solutions

<u>model</u> : $(n_0, n_1, n_2, n_3) \sim MULT(n, \pi_0, \pi_1, \pi_2, \pi_3)$

$(\underline{\pi} = \underline{\pi}_0(\theta)$, where

$H_0 : \pi_{00}(\theta) = 1-\theta, \quad \pi_{10}(\theta) = \theta(1-\theta), \quad \pi_{20}(\theta) = \theta^2(1-\theta), \quad \pi_{30}(\theta) = \theta^3$

(1.) $L(\theta) = n_0 \log(1-\theta) + n_1 \log\left(\theta(1-\theta)\right) + n_2 \log\left(\theta^2(1-\theta)\right) + n_3 \log(\theta^3)$

$$= (n_1 + 2n_2 + 3n_3)\cdot \log\theta + (n_0 + n_1 + n_2)\cdot\log(1-\theta)$$

$$\frac{dL}{d\theta} = \frac{n_1 + 2n_2 + 3n_3}{\theta} - \frac{n_0 + n_1 + n_2}{1-\theta}.$$ Solve $\frac{dL}{d\theta} = 0$ to get

$$\boxed{\hat{\theta} = \frac{n_1 + 2n_2 + 3n_3}{n_0 + 2n_1 + 3n_2 + 3n_3}}$$

$\hat{\theta} = \frac{\#\text{ infected}}{\#\text{ at risk}}$

(2.) <u>data</u> : $n_0 = 63, \; n_1 = 63, \; n_2 = 25, \; n_3 = 5 \quad (n = 156)$

$\Rightarrow \quad \underline{\underline{\hat{\theta} = 0.4587814}}$

$\underline{\pi}_0(\hat{\theta}) = (.5412, .2483, .1139, .0966)$

3. $\hat{\underline{m}} = n\,\underline{\pi}_0(\hat{\theta}) = (84.43,\ 38.73,\ 17.77,\ 15.06)$

(a) $G^2 = 2\sum_j n_j \log\left(\frac{n_j}{\hat{m}_j}\right) = \underline{\underline{30.43}}$, $df = c-1-t = 4-1-1 = 2$

p-value $= P(\chi^2_2 > 30.43) < .001$

(b) The data provides very strong evidence against the constant infection rate model.

4. $\hat{\pi}_j = \frac{n_j}{n}$, $\underline{\underline{\hat{\underline{\pi}} = (.4038,\ .4038,\ .16025,\ .03205)}}$

(a)

$$\left(\hat{\underline{\pi}} = \left(\frac{63}{156}, \frac{63}{156}, \frac{25}{156}, \frac{5}{156}\right) \right)$$

(b) $\hat{P}(\text{primary}) = \frac{63+25+5}{156} = .596$

$\hat{P}(\text{secondary} \mid \text{primary}) = \frac{25+5}{63+25+5} = .3226$

$\hat{P}(\text{tertiary} \mid \text{secondary}) = \frac{5}{25+5} = .1667$

(c) The data is compatible with a decreasing infection rate model.

```
> obs = c(63,63,25,5)
> n = sum(obs)
> theta = (3*obs[4]+2*obs[3]+obs[2]) / (3*obs[4]+3*obs[3]+2*obs[2]+obs[1])
> theta
[1] 0.4587814
> t = 1
> c = length(obs)
> pi = rep(NA,c)
> pi[1] = 1-theta
> pi[2] = theta*(1-theta)
> pi[3] = theta^2*(1-theta)
> pi[4] = theta^3
> print(pi,digits=3)
[1] 0.5412 0.2483 0.1139 0.0966


> m = n*pi
> m
[1] 84.43011 38.73496 17.77088 15.06406
>
> G2 = 2*sum(obs*log(obs/m))
> df = c-1-t
> p.value = pchisq(G2,c-1-t,lower.tail = FALSE)
>
> print(G2)
[1] 30.43096
> print(df)
[1] 2
> print(p.value)
[1] 2.466041e-07



> pi.full = obs / n
> pi.full
[1] 0.40384615 0.40384615 0.16025641 0.03205128

> prob.primary = pi.full[2]+pi.full[3]+pi.full[4]
> prob.primary
[1] 0.5961538

> prob.secondary = (pi.full[3]+pi.full[4])/(pi.full[2]+pi.full[3]+pi.full[4])
> prob.secondary
[1] 0.3225806


> prob.tertiary = pi.full[4] / (pi.full[3]+pi.full[4])
> prob.tertiary
[1] 0.1666667
```