

Exercise5

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

3/1/2020

```
transplant2 <- read.table("./transplant.txt", quote="\"", comment.char="")
colnames(transplant2)<-c("time","type","s")
```

(1)

Because the failure rate doesn't dependent on the number of individuals.

Let the probability of individuals that die/relapse be $p_A^{(k)} = p_B^{(k)} = p$ (under H_0).

So $P(y^{(k)} = m, n_d^{(k)} = m_d | n_A^{(k)} = m_A, n_B^{(k)} = m_B) = \binom{m_A}{m} p^m (1-p)^{m_A-m} \binom{m_B}{m_d-m} p^{m_d-m} (1-p)^{m_B-m_d+m} = \binom{m_A}{m} \binom{m_B}{m_d-m} p^{m_d} (1-p)^{m_A+m_B-m_d}$

$P(n_d^{(k)} = m_d | n_A^{(k)} = m_A, n_B^{(k)} = m_B) = \binom{m_A+m_B}{m_d} p^{m_d} (1-p)^{m_A+m_B-m_d}$

So $P(y^{(k)} = m | n_d^{(k)} = m_d, n_A^{(k)} = m_A, n_B^{(k)} = m_B) = \frac{P(y^{(k)}=m, n_d^{(k)}=m_d | n_A^{(k)}=m_A, n_B^{(k)}=m_B)}{P(n_d^{(k)}=m_d | n_A^{(k)}=m_A, n_B^{(k)}=m_B)} = \frac{\binom{m_A}{m} \binom{m_B}{m_d-m}}{\binom{m_A+m_B}{m_d}}$

It's HyperGeometric $\left(n_A^{(k)} + n_B^{(k)}, n_A^{(k)}, n_d^{(k)}\right)$.

(2)

$$P = \frac{\binom{n_A}{y} \binom{n_B}{n_d-y}}{\binom{n}{n_d}}$$

Let $Y_i = \begin{cases} 1 & \text{if the } i \text{ th selection is under treatment A} \\ 0 & \text{otherwise} \end{cases}$

So $P(Y_i = 1) = \frac{n_A}{n}$, $P(Y_i = 1, Y_j = 1) = \frac{n_A}{n} \frac{n_A-1}{n-1}$

$$E^{(k)} = E(\sum Y_i) = n_d EY = \frac{n_A^{(k)} n_d^{(k)}}{n^{(k)}}$$

$$Var^{(k)} = Var(\sum Y_i) = n_d VarY + n_d(n_d-1)Cov(Y_i, Y_j)$$

$$VarY = \frac{n_A n_B}{n^2}, Cov(Y_i, Y_j) = EY_i Y_j - EY_i EY_j = P(Y_i = 1, Y_j = 1) - \left(\frac{n_A}{n}\right)^2 = -\frac{n_A n_B}{n^2(n-1)}$$

$$\text{Then } Var^{(k)} = \frac{n_A n_B n_d}{n^2} - \frac{n_A n_B n_d (n_d-1)}{n^2(n-1)} = \frac{n_A^{(k)} n_B^{(k)} n_d^{(k)} n_s^{(k)}}{(n^{(k)})^2 (n^{(k)}-1)}$$

(3)

Let condition $\left(n_A^{(k)}, n_B^{(k)}, n_d^{(k)}\right)$ be C_i .

$$\text{Var}[y^{(k)} - E^{(k)}] = \text{Var}[\mathbb{E}[y^{(k)} - E^{(k)} | C_i]] + \mathbb{E}[\text{Var}[y^{(k)} - E^{(k)} | C_i]]$$

As we showed in (2), $\mathbb{E}[y^{(k)} - E^{(k)} | C_i] = 0$, $\text{Var}[y^{(k)} - E^{(k)} | C_i] = V^{(k)}$.

$$\text{So } \text{Var}[y^{(k)} - E^{(k)}] = E[V^{(k)}]$$

(4)

$$\text{Var} \left[\sum_{k=1}^K (y^{(k)} - E^{(k)}) \right] = \sum_{k=1}^K \text{Var} [y^{(k)} - E^{(k)}] + 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K \text{Cov} [y^{(i)} - E^{(i)}, y^{(j)} - E^{(j)}]$$

Let the condition $(n_A^{(i)}, n_B^{(i)}, n_d^{(i)}, n_A^{(j)}, n_B^{(j)}, n_d^{(j)})$ be C_{ij}

With Law of Total Variance, we can get

$$\begin{aligned} & \text{Cov} [y^{(i)} - E^{(i)}, y^{(j)} - E^{(j)}] \\ &= \text{Cov} [\mathbb{E}[y^{(i)} - E^{(i)} | C_{ij}], \mathbb{E}[y^{(j)} - E^{(j)} | C_{ij}]] + \mathbb{E}[\text{Cov}[y^{(i)} - E^{(i)}, y^{(j)} - E^{(j)} | C_{ij}]] \end{aligned}$$

$$\mathbb{E}[y^{(i)} - E^{(i)} | C_{ij}] = \mathbb{E}[y^{(i)} - E^{(i)} | C_i] = 0$$

$$\text{So Cov} [\mathbb{E}[y^{(i)} - E^{(i)} | C_{ij}], \mathbb{E}[y^{(j)} - E^{(j)} | C_{ij}]] = \text{Cov}(0, 0) = 0$$

$$\text{Cov}[y^{(i)} - E^{(i)}, y^{(j)} - E^{(j)} | C_{ij}] = \text{Cov}[y^{(i)}, y^{(j)} | C_{ij}]$$

Under the condition C_{ij} , $y^{(i)}, y^{(j)}$ are two independent hyperGeometric distributions, so the covariance of them is 0.

$$\text{Then Cov} [y^{(i)} - E^{(i)}, y^{(j)} - E^{(j)}] = 0 \quad \forall i \neq j$$

$$\text{So Var} \left[\sum_{k=1}^K (y^{(k)} - E^{(k)}) \right] = \sum_{k=1}^K \text{Var} [y^{(k)} - E^{(k)}] = \sum_{k=1}^K \mathbb{E} [V^{(k)}]$$

(5)

```
library(dplyr)
```

```
n <- nrow(transplant2)
k <- nrow(transplant2 %>% filter(s==1))
na <- nrow(transplant2 %>% filter(type==1))
nb <- nrow(transplant2 %>% filter(type==2))
df <- transplant2[order(transplant2$time),]
df <- df %>% filter(s==1)
y = c()
e = c()
v = c()
for (i in 1:k) {

  e[i] <- na/(na+nb)
  v[i] <- na*nb*(na+nb-1)/(na+nb)^2/(na+nb-1)

  if (df$type[i]==1) {
    y[i] <- 1
    na <- na - 1
  } else {
    y[i] <- 0
    nb <- nb - 1
  }
}

z <- sum(y-e)/sqrt(sum(v))
print(pnorm(z))
```

```
## [1] 0.2828261
```

We can accept H_0 at 0.05 significant level.