414HW4

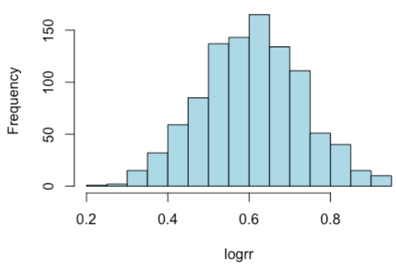
Ningyuan Wang

10/6/2020

1

a. The histogram of the log relative risk was shown below, and the estimated standard error is 0.1236. 95% CI for the RR is (1.4268, 2.3160)

```
set.seed(456)
dat = c(189, 104, 10845, 10933)
myoasp = matrix(dat, nrow = 2, ncol = 2)
colnames(myoasp) = c("Yes", "No")
rownames(myoasp) = c("Placebo", "Aspirin")
simdat = rmultinom(1000, size = 22071,
                   prob = c(189, 104, 10845, 10933)/22071)
logthfun = function(thta){
  log((thta[1]/(thta[1]+thta[3])) / (thta[2]/(thta[2]+thta[4])))}
logrr = apply(simdat, 2, logthfun)
hist(logrr, col="lightblue", breaks = 15, main = "")
```




```
round(sd(logrr), 4)

## [1] 0.1236

# 95% CI
dat_log = log((189/(189+10845)) / (104/(104+10933)))#0.6071993
exp(dat_log - 1.96*sd(logrr))

## [1] 1.426781

exp(dat_log + 1.96*sd(logrr))

## [1] 2.315986
```

b. The estimated standard error is 0.1213, which is close to the result in part a

```r
pihat = cbind(c(189, 10845, 104, 10933)/22071)
sigma = diag(c(pihat)) - pihat %*% t(pihat)
temp = c(-1/(pihat[1] + pihat[2]), 1/(pihat[3] + pihat[4]))
B = rbind(c(temp[1]+(1/pihat[1]), temp[1], temp[2]-(1/pihat[3]), temp[2]))
sqrt(B %*% sigma %*% t(B)/22071)

##            [,1]
## [1,] 0.1213473
```

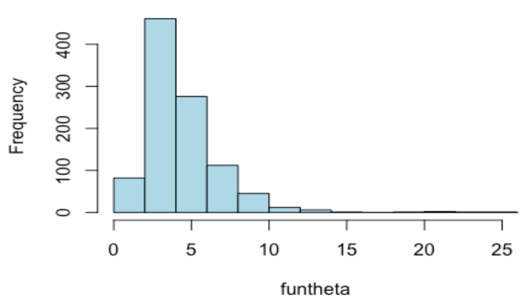
2

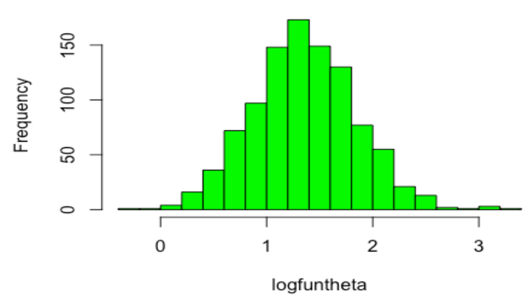
a. The histogram of the odds ratio is shown below, and the standard error of the sample is 2.4569.

b. The standard error of the log odds ratio samples is 0.4893. The log odds ratio follows a normal distribution. The log odds ratio is 1.3218 (please check the hand-written part below) and 95% CI of theta is (1.4391, 9.7726)

c. The estimated standard error of log odds ratio calculated by hand is 0.4726, and the 95% CI is (1.4815, 9.4697), which is pretty close to the result in part b. Please check the hand-written part below. The small difference is probably because of the sample size.

```r
set.seed(456)
dat2 = c(30, 10, 20, 25)
myoasp2 = matrix(dat2, nrow = 2, ncol = 2)
colnames(myoasp2) = c("Yes", "No")
rownames(myoasp2) = c("Yes", "No")
simdat2 = rmultinom(1000, size = 85,
                    prob = c(30, 20, 10, 25)/85) # row-based
fun = function(thta){
  (thta[1] * thta[4])/ (thta[2]* thta[3])}
funtheta = apply(simdat2, 2, fun)
hist(funtheta, col="lightblue", breaks = 15, main = "")
```






```r
round(sd(funtheta), 4)

## [1] 2.4569
```

```r
#b.
logfun = function(thta){
  log((thta[1] * thta[4])/ (thta[2]*(thta[3])))}
logfuntheta = apply(simdat2, 2, logfun)
hist(logfuntheta, col="green", breaks = 15, main = "")

round(sd(logfuntheta), 4)

## [1] 0.4893

dat_log2 = 1.3218
exp(dat_log2 - 1.96*sd(logfuntheta))

## [1] 1.437172

exp(dat_log2 + 1.96*sd(logfuntheta))

## [1] 9.785703
```

2b. $\log \hat{\theta} = \log\left(\frac{n_{11} n_{22}}{n_{12} n_{21}}\right) = \log\left(\frac{30 \cdot 25}{20 \cdot 10}\right) = \log(3.75)$
$= 1.3218$

2c. $\hat{se}(\log \hat{\theta}) = \sqrt{\frac{1}{30} + \frac{1}{20} + \frac{1}{10} + \frac{1}{25}}$
$= 0.4726$

95% CI:

$e^{1.3218 \pm 1.96 \cdot 0.4726} = e^{(0.3955,\ 2.2481)}$

$= (1.4851,\ 9.4697)$

3a. $\sum_{i=1}^{2}\sum_{j=1}^{2}(n_{ij} - \hat{u}_{ij}) = (n_{11} - \hat{u}_{11}) + (n_{12} - \hat{u}_{12}) + (n_{21} - \hat{u}_{21}) + (n_{22} - \hat{u}_{22})$

$= n_{11} - \frac{n_{1+} n_{+1}}{n} + n_{12} - \frac{n_{1+} n_{+2}}{n} + n_{21} - \frac{n_{2+} n_{+1}}{n} + n_{22} - \frac{n_{2+} n_{+2}}{n}$

$= \left(n_{1+} - \frac{n_{1+} n_{+1}}{n} - \frac{n_{1+} n_{+2}}{n}\right) + \left(n_{2+} - \frac{n_{2+} n_{+1}}{n} - \frac{n_{2+} n_{+2}}{n}\right)$

$= n_{1+}\left[1 - \frac{n_{+1} + n_{+2}}{n}\right] + n_{2+}\left[1 - \frac{n_{+1} + n_{+2}}{n}\right]$

$= 0$

3b. $2\sum_{i=1}^{2}\sum_{j=1}^{2} f(n_{ij}) \approx 2\sum_{i=1}^{2}\sum_{j=1}^{2} f(\hat{u}_{ij}) + (n_{ij} - \hat{u}_{ij}) f'(u_{ij}) + \frac{1}{2}(n_{ij} - u_{ij})^2 f''(u_{ij})$

find out $f(\hat{u}_{ij})$, $f'(\hat{u}_{ij})$, and $f''(\hat{u}_{ij})$:

since $f(n_{ij}) = n_{ij} \log(n_{ij} \mid \hat{u}_{ij}) = n_{ij} \log(n_{ij}) - n_{ij} \log(\hat{u}_{ij})$

$$f'(n_{ij}) = \log(n_{ij}) + 1 - \log(\hat{n}_{ij})$$

$$f''(n_{ij}) = \frac{1}{n_{ij}}$$

So, $f(\hat{n}_{ij}) = 0$

$$f'(\hat{n}_{ij}) = 1$$

$$f''(\hat{n}_{ij}) = \frac{1}{\hat{n}_{ij}}$$

$$= 2\sum_{i=1}^{2}\sum_{j=1}^{2}\left[0 + \overbrace{(n_{ij} - \hat{n}_{ij})}^{=0} + \frac{1}{2}\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}\right]$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = \chi^2$$

pearson chi-squared