STAT931 Assignment 1

Branden Lee 20877653

28/09/2020

Question 2

a) $\widehat{RR} = \frac{\widehat{P(D|E)}}{\widehat{P(D|E^c)}} = \frac{6/70}{60/1571} = 2.2442857$

Denoting $N_.975$ to be the .975 quantile of the standard normal distribution, an approximate 95% confidence interval for $\widehat{RR}$ is

$$\widehat{RR} \cdot exp\left(\pm N_.975\sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}\right) = (1.0040671, 5.016416)$$

We can say with 95% confidence that binge drinkers are somewhere between roughly 1 to 5 times more likely to succumb to cardiovascular death than non-binge drinkers.

b) $\widehat{OR} = \frac{\widehat{P(D|E)}/\widehat{P(D^c|E)}}{\widehat{P(D|E^c)}/\widehat{P(D^c|E^c)}} = \frac{ad}{bc} = \frac{6 \cdot 1511}{64 \cdot 60} = 2.3609375$

An approximate 95% confidence interval for $\widehat{OR}$ is

$$\widehat{OR} \cdot exp\left(\pm N_.975\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right) = (0.9835065, 5.6675026)$$

We can say with 95% confidence that the odds of succumbing to cardiovascular death for binge drinkers versus non binge drinkers is somewhere between roughly .98 and 5.67. The similarity in the magnitude of the estimated odds ratio and the estimated relative risk is to be expected, as the estimated ratio $\frac{\widehat{P(D^c|E^c)}}{\widehat{P(D|E^c)}} = \frac{1511/1571}{64/70} = 1.0519772 \approx 1$

c) $\widehat{ER} = \widehat{P(D|E)} - \widehat{P(D|E^c)} = \frac{6}{70} - \frac{60}{1571} = 0.0475221$

An approximate 95% confidence interval for $\widehat{ER}$ is

$$\widehat{ER} \pm N_.975\sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}} = (-0.0187385,\ 0.1137826)$$

We can say with 95% confidence that the risk of succumbing to cardiovascular death as a binge drinker is somewhere between roughly -2% to 11% more than the risk of a non-binge drinker succumbing.

d) $\widehat{AR} = \frac{\widehat{P(E)}(\widehat{RR}-1)}{\widehat{P(E)}(\widehat{RR}-1)+1} = 0.0504022$

An approximate 95% confidence interval for $\widehat{AR}$ is

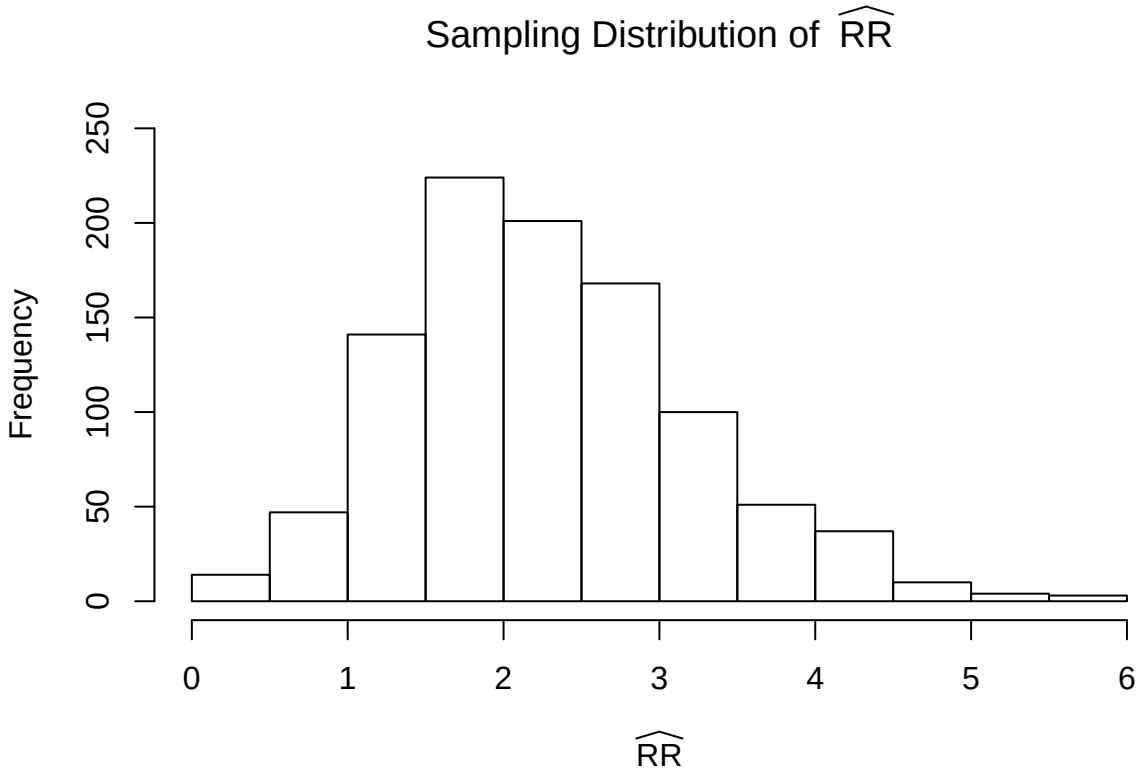
$$1 - (1 - \widehat{AR}) \cdot exp\left(\pm N_{.975}\sqrt{\frac{b + \widehat{AR}(a+d)}{nc}}\right) = (-0.0225618, 0.1181599)$$

We can say with 95% confidence that binge drinking can be attributed to somewhere between roughly -2% to 12% of cardiovascular deaths.

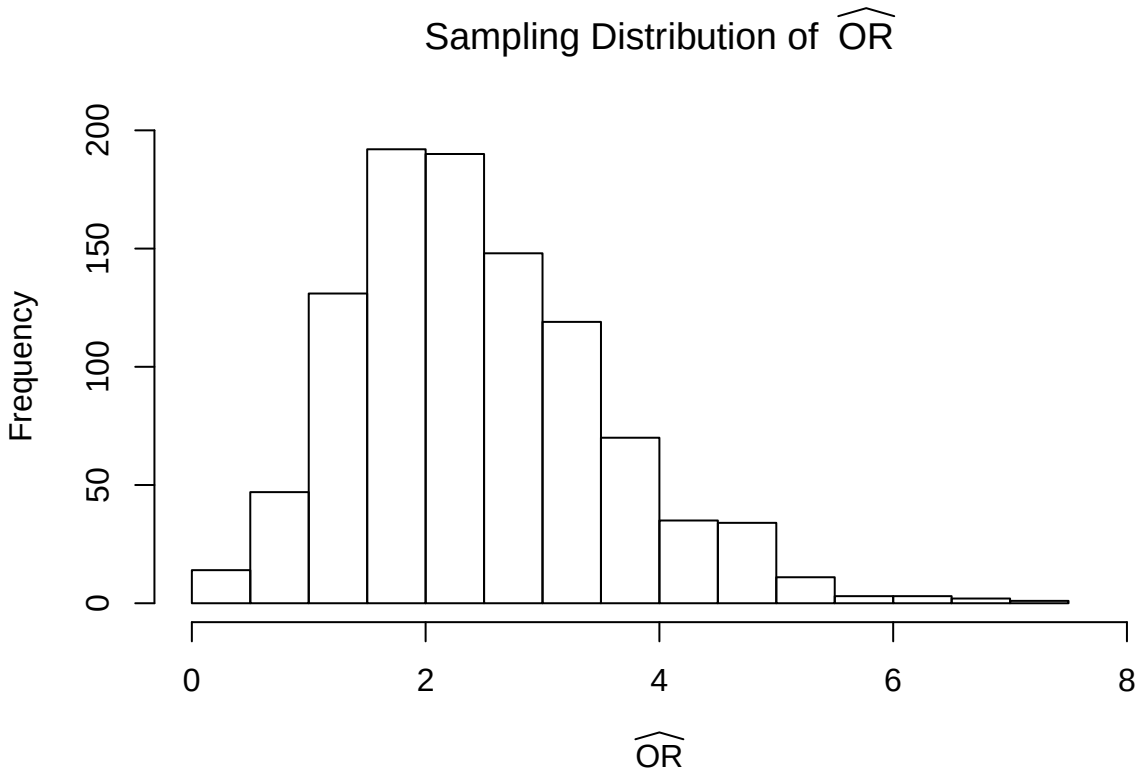
e) We can be sure that the risk factor (binge drinking) occurred before the outcome of interest (cardiovascular death). The confidence intervals for the various risk measures also suggest binge drinking is associated with the risk of cardiovascular death. However, because the data only explored only one possible risk factor in a specific demographic group and did not investigate any potential confounding risk factors (ex possible relationship between age/gender and cardiovascular health), it would be hasty to conclude from this data that binge drinking in general causes an increase in the risk of cardiovascular death.

f)

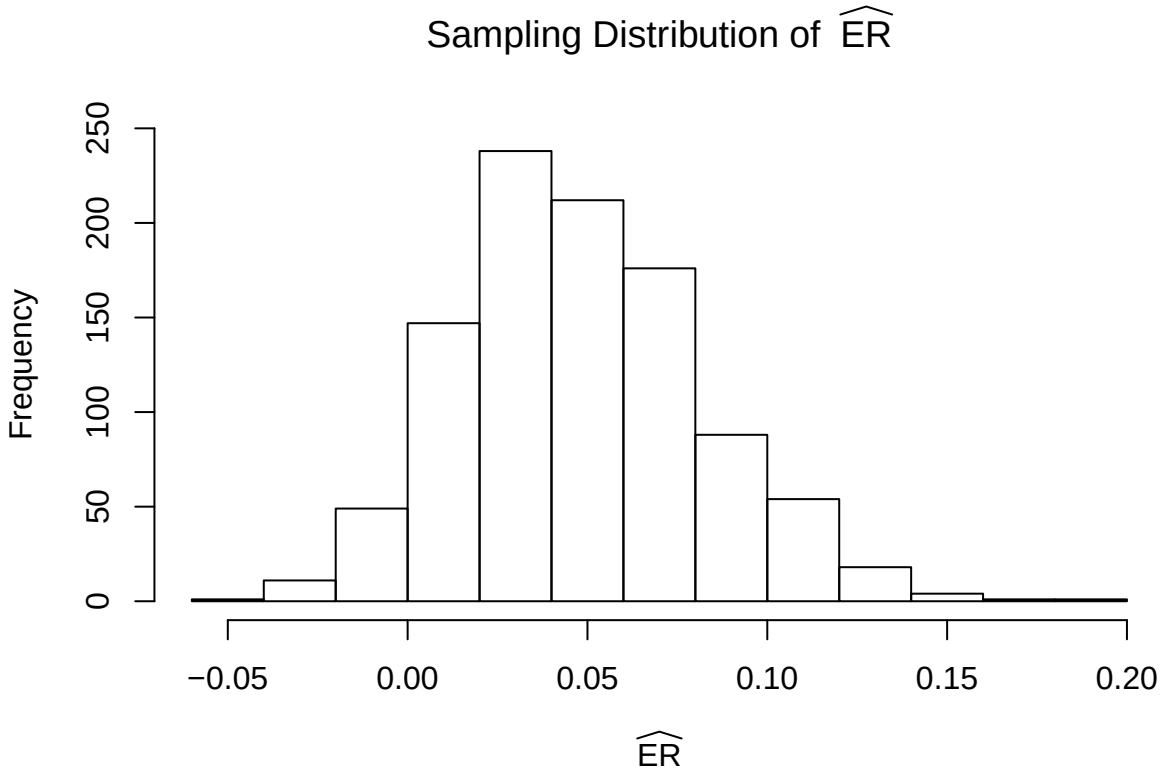
```r
set.seed(0)

k <- 1000
a <- 6
c <- 60
n1 <- 70
n2 <- 1571

sample_disease_exposed <- rbinom(k, n1, a/n1)
sample_disease_not_exposed <- rbinom(k, n2, c/n2)

relative_risk <- function(p1, p2) {p1 / p2}
odds_ratio <- function(p1, p2) {(p1 / (1 - p1)) / (p2 / (1 - p2))}
excess_risk <- function(p1, p2) {p1 - p2}

rr <- relative_risk(sample_disease_exposed / n1, sample_disease_not_exposed / n2)
or <- odds_ratio(sample_disease_exposed / n1, sample_disease_not_exposed / n2)
er <- excess_risk(sample_disease_exposed / n1, sample_disease_not_exposed / n2)

hist(rr, xlab=bquote(widehat(RR)), ylim=c(0,250),
     main=bquote('Sampling Distribution of ' ~ widehat(RR)))
```




```
hist(or, xlab=bquote(widehat(OR)), ylim=c(0,200), xlim=c(0,8),
     main=bquote('Sampling Distribution of ' ~ widehat(OR)))
```




```
hist(er, xlab=bquote(widehat(ER)), ylim=c(0,250),
     main=bquote('Sampling Distribution of ' ~ widehat(ER)))
```




```
quantile(rr, c(.025, .975))
```

```
##      2.5%     97.5%
## 0.7239631 4.2980275
```

```
quantile(or, c(.025, .975))
```

```
##      2.5%     97.5%
## 0.7158444 4.8686692
```

```
quantile(er, c(.025, .975))
```

```
##        2.5%       97.5%
## -0.01089388  0.11896654
```

The sampling distributions of the risks look normally distributed. Comparing the initial confidence intervals with the new ones, the new intervals for relative risk and odds ratio have smaller endpoints, while the new interval for excess risk has been shifted to the right compared to the initial interval.