epidemiology measures

SURV686

Example from class:

Table: Data for Example 1 and 2

Smoker	Stroke Yes	Stroke No	Total
Yes	171	3,264	3,435
No	117	4,320	4,437
Total	288	7,584	7,872

Please calculate incidence among smokers, non-smokers, and the relative risk and odds ratio for smokers compared to non-smokers.

We enter the data as separate vectors to facilitate calculations.

```
library(tidyverse)
library(epiR)
a <- 171
b <- 3264
c <- 117
d <- 4320

# total
t <- sum(a, b, c, d)
```

Calculate incidence among smokers and non-smokers using;

$$I_{Exposed} = \frac{A}{A+B}$$

$$I_{Un-exposed} = \frac{C}{C+D}$$

```
ir_exp <- a / (a+b) ; ir_exp
```

```
[1] 0.04978166
```

```
ir_unexp <- c / (c+d) ; ir_unexp
```

```
[1] 0.02636917
```

- About 5% of smokers had a stroke, almost twice as many as the non-smokers.

Calculate relative risk:

$$RR = \frac{I_{Exposed}}{I_{Un-exposed}}$$

$$

$$

```
RR <- round(ir_exp / ir_unexp, 4)
```

- Smokers are 1.88 times more at risk of having a stroke compared to non-smokers.

Calculate Odds Ratio

$$\hat{OR} = \frac{AD}{BC}$$

```
OR <- (a*d) / (b*c)
```

- Our odds ratio is close to our relative risk estimate.

Calculate Variances for OR and RR, to estimate 95% CI.

$$Var(\ln \hat{R}) = \frac{b}{a(a+b)} + \frac{d}{c(c+d)}$$

$$Var(\ln \hat{O}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

```
var_RR <- (b / (a*(a+b))) + (d / (c*(c+d))) ; var_RR
```

```
[1] 0.01387846
```

```
var_OR <- (1/a) + (1/b) + (1/c) + (1/d) ; var_OR
```

```
[1] 0.01493282
```

```
lower <- round(exp(log(RR) - 1.96 * sqrt(var_RR)) , 4)
upper <- round(exp(log(RR) + 1.96 * sqrt(var_RR)) , 4)

print(str_c("Printing 95% CI for relative risk . . . ", RR,
            " [", lower, ", ",  upper, "]"))
```

```
[1] "Printing 95% CI for relative risk . . . 1.8879 [1.4986, 2.3783]"
```

```
lower <- round(exp(log(OR) - 1.96 * sqrt(var_OR)), 4)
upper <- round(exp(log(OR) + 1.96 * sqrt(var_OR)), 4)

print(str_c("Printing 95% CI for odds ratio. . . ", RR,
            " [", lower, ", ", upper, "]"))
```

```
[1] "Printing 95% CI for odds ratio. . . 1.8879 [1.5224, 2.4579]"
```

- The intervals do not contains 1, so they are statistically significant and reject the Null hypothesis that the odds are equal across smokers and non-smokers.

Prospective study:

```r
a <- 23
b <- 125
c <- 13
d <- 150
t <- sum(a, b, c, d)
# check attrition
attrition <- sum(a/t, b/t, c/t, d/t) ; attrition
```

```
[1] 1
```

Calculate relative risk, and odds ratio from formula from before.

$$\hat{R} = \frac{a(c+d)}{c(a+b)}$$

```r
R_hat <- (a*(c+d)) / (c*(a+b)) ; R_hat
```

```
[1] 1.948545
```

```r
OR <- (a*d)/(b*c) ;  OR
```

```
[1] 2.123077
```

- The odds ratio is slightly larger than the relative risk.

Calculate the attributed risk percent in the exposed group.

$$A\hat{R}P_{Exposed}\frac{\hat{R}-1}{\hat{R}}$$

```r
ARP_exposed <-  (R_hat - 1) / R_hat ; ARP_exposed # fraction
```

```
[1] 0.4867965
```

- The amount of disease incidence is 49% which can be attributed to an exposure in a prospective study.
- This is also refereed to ARP, attributable risk percent when converting to a percent.

The attributable risk (ratio) of the exposed is calculated as incidence of the exposed - incidence of the non-exposed.

```r
ir_exp <- a / (a+b)
ir_unexp <- c / (c+d)
AR_exposed <- ir_exp - ir_unexp ; AR_exposed
```

```
[1] 0.0756508
```

To calculate 95% CI we would need to calculate the standard error as:

$$SE_{AR} = \sqrt{\frac{a+c}{N}(1 - \frac{a+c}{N})(\frac{1}{a+b} + \frac{1}{c+d})}$$

```r
SE_AR <- sqrt(
  ((a+c)/t)*(1-((a+c)/t )) * ( (1/(a+b)) + (1/(c+d)) )
)

(AR_exposed - 1.97*SE_AR) * 100 # lower
```

```
[1] 0.4089381
```

```r
(AR_exposed + 1.97*SE_AR) * 100 # Upper
```

```
[1] 14.72122
```

When we use the `epiR::epi.2by2()`we can see these results under "Attrib prev in the exposed."

Calculate the attributable risk in the population

$$\hat{A}_{pop} = \frac{ad - bc}{(a+c)(c+d)}$$

```r
# reduction in incidence in the population that would
# occur in the absence of the risk factor
A_hat_pop <- (a*d - b*c) / ((a+c)* (c+d)) ; A_hat_pop # fraction
```

```
[1] 0.3110089
```

- This is the reduction in incidence of 31% if the whole population were unexposed, comparing with actual exposure.
- This is known as PARP, population attributable risk when converting to percent.

Calculate the variance of the attributable risk in the population.

$$V(\ln(1-\hat{A}_{pop}) = \frac{b + \hat{A}_{pop}(a+d)}{tc}$$

```
V <- (b + A_hat_pop* (a+d)) / (t*c) ;  V
```

```
[1] 0.04422571
```

In a retrospective study we use the OR instead of or relative risk, and calculate the attributable risk in the exposed group and population.

```
R_hat <- (a*d)/(b*c) ; R_hat
```

```
[1] 2.123077
```

```
A_exposed <-  (R_hat - 1) / R_hat ; A_exposed # fraction
```

```
[1] 0.5289855
```

For the population we change the denominator.

$$\hat{A}_{pop} = \frac{ad - bc}{d(a+c)}$$

```
A_hat_pop <- (a*d - b*c) / ( d * (a+c)) ; A_hat_pop # fraction
```

```
[1] 0.337963
```

The variance is also computed differently

$$V(\ln(1-\hat{A}_{pop}) = \frac{a}{c(a+c)} + \frac{b}{d(b+d)}$$

```r
V <- (a / (c * (a + c))) +
  (b / (d*(b+d)));  V
```

```
[1] 0.0521756
```

We can now calculate 95% CI.

$$LCL = 1 - \exp\left(\ln(1 - \hat{A}_{pop}) + 1.96\sqrt{V(\ln(1 - \hat{A}_{pop}))}\right)$$

$$UCL = 1 - \exp\left(\ln(1 - \hat{A}_{pop}) - 1.96\sqrt{V(\ln(1 - \hat{A}_{pop}))}\right)$$

```r
LCL <- 1 - exp( A_hat_pop + 1.96 * sqrt(V) ) ; LCL
```

```
[1] -1.193867
```

```r
UCL <- 1 - exp( A_hat_pop - 1.96 * sqrt(V) ) ; UCL
```

```
[1] 0.1039328
```

We can also use the r package epiR

```r
# create table for later
bp <- matrix(c(a, b, c, d), ncol = 2, byrow = TRUE)
epi.2by2(bp, method="cross.sectional")
```

```
             Outcome +    Outcome -      Total              Prev risk *
Exposed +           23          125        148    15.54 (10.11 to 22.40)
Exposed -           13          150        163      7.98 (4.31 to 13.25)
Total               36          275        311     11.58 (8.24 to 15.66)

Point estimates and 95% CIs:
-------------------------------------------------------------------
Prev risk ratio                                  1.95 (1.02, 3.71)
Prev odds ratio                                  2.12 (1.03, 4.36)
Attrib prev in the exposed *                     7.57 (0.40, 14.73)
Attrib fraction in the exposed (%)              48.68 (2.41, 73.01)
Attrib prev in the population *                  3.60 (-1.87, 9.07)
Attrib fraction in the population (%)           31.10 (-4.04, 54.37)
-------------------------------------------------------------------
Uncorrected chi2 test that OR = 1: chi2(1) = 4.337 Pr>chi2 = 0.037
Fisher exact test that OR = 1: Pr>chi2 = 0.050
 Wald confidence limits
 CI: confidence interval
 * Outcomes per 100 population units
```