How Small Cell Sizes -> Large OR's

Simulating Some Hypothetical Data

Andy Grogan-Kaylor

2024-05-30

Table of contents

1 call library

```
library(dplyr) # data wrangling
```

2 set up the initial data

Note that row 1 has a small cell size / very infrequent event.

3 make some calculations

Table 1: Table of Data

```
condition <- c("A", "B", "C")

total_events <- c(100, 100, 100)

event_happened <- c(1, 10, 19)

mytable <- data.frame(condition,
                      total_events,
                      event_happened)

mytable # replay
```

```
  condition total_events event_happened
1         A          100              1
2         B          100             10
3         C          100             19
```

Table 2: Table of Calculations

```
mytable %>%
  mutate(risk = event_happened / total_events) %>%
  mutate(risk_difference = risk - lag(risk)) %>%
  mutate(event_didnt_happen = total_events - event_happened) %>%
  mutate(odds = event_happened / event_didnt_happen) %>%
  mutate(odds_ratio = odds / lag(odds)) %>%
  mutate(comment = case_when(row_number() == 1 ~ "odds ratio not defined",
                             row_number() == 2 ~"OR 2 vs. 1 is large",
                             row_number() == 3 ~ "OR 3 vs. 2 is reasonable"))
```

```
  condition total_events event_happened risk risk_difference event_didnt_happen
1         A          100              1 0.01              NA                 99
2         B          100             10 0.10            0.09                 90
3         C          100             19 0.19            0.09                 81
        odds odds_ratio                  comment
1 0.01010101         NA   odds ratio not defined
2 0.11111111  11.000000      OR 2 vs. 1 is large
3 0.23456790   2.111111 OR 3 vs. 2 is reasonable
```