Patel's Dataset - Conditional Probability Lab

Harsh Patel

2024-09-20

```r
gender = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.5, 0.5))
smoking = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.6, 0.4))

number.draws = 2
replicates = 10

successes = vector("numeric", replicates)

for(k in 1:replicates){

  draw = sample(1:1659, size = number.draws, replace = FALSE)

  if(gender[draw[1]] == 1 & smoking[draw[1]] == 1 & gender[draw[2]] == 1 & smoking[draw[2]] == 1){
successes[k] = 1
  }

}

successes
```

```
##  [1] 0 0 0 0 1 0 0 0 0 0
```

```r
table(successes)
```

```
## successes
## 0 1
## 9 1
```

```r
which(successes == 1)
```

```
## [1] 5
```

```r
gender = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.5, 0.5))
smoking = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.6, 0.4))

number.draws = 2
replicates = 10000

successes = vector("numeric", replicates)
```

```r
for(k in 1:replicates){

  draw = sample(1:1659, size = number.draws, replace = FALSE)

  if(gender[draw[1]] == 1 & smoking[draw[1]] == 1 & gender[draw[2]] == 1 & smoking[draw[2]] == 1){
successes[k] = 1

  }

}

table(successes)
```

```
## successes
##    0    1
## 9651  349
```

```r
sum(successes)/replicates
```

```
## [1] 0.0349
```

```r
gender = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.5, 0.5))
smoking = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.6, 0.4))

number.draws = 2
replicates = 10000

successes = vector("numeric", replicates)

for(k in 1:replicates){

  draw = sample(1:1659, size = number.draws, replace = FALSE)

  if(gender[draw[1]] == 1 & smoking[draw[1]] == 1 & gender[draw[2]] == 0 & smoking[draw[2]] == 0){
    successes[k] = 1

  }

  if(gender[draw[1]] == 0 & smoking[draw[1]] == 0 & gender[draw[2]] == 1 & smoking[draw[2]] == 1){
    successes[k] = 1

  }

}

table(successes)
```

```
## successes
##    0    1
## 8882 1118
```

```r
sum(successes)/replicates
```

```
## [1] 0.1118
```

```r
gender = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.5, 0.5))
smoking = sample(c(0, 1), size = 1659, replace = TRUE, prob = c(0.6, 0.4))

number.draws = 2
replicates = 10000

successes = vector("numeric", replicates)

for(k in 1:replicates){

  draw = sample(1:1659, size = number.draws, replace = FALSE)

  if(gender[draw[1]] == 1 & smoking[draw[1]] == 1 & gender[draw[2]] == 0 & smoking[draw[2]] == 0){
    successes[k] = 1

  }

  if(gender[draw[1]] == 0 & smoking[draw[1]] == 0 & gender[draw[2]] == 1 & smoking[draw[2]] == 1){
    successes[k] = 1

  }

}

table(successes)
```

```
## successes
##    0    1
## 8849 1151
```

```r
sum(successes)/replicates
```

```
## [1] 0.1151
```

```r
p.female = 0.50
p.smoking.if.female = 0.40
p.smoking.if.male = 0.60
population.size = 1659

gender = sample(c(0, 1), size = population.size, prob = c(1 - p.female, p.female), replace = TRUE)
smoking = vector("numeric", population.size)

for (k in 1:population.size){

  if (gender[k] == 0) {
    smoking[k] = sample(c(0, 1), prob = c(1 - p.smoking.if.male, p.smoking.if.male), size = 1)
```

```
  }

  if (gender[k] == 1) {
    smoking[k] = sample(c(0, 1), prob = c(1 - p.smoking.if.female, p.smoking.if.female), size = 1)

  }

}

addmargins(table(gender, smoking))
```

```
##       smoking
## gender   0    1  Sum
##    0   339  514  853
##    1   488  318  806
##    Sum 827  832 1659
```

```
sum(smoking == 1 & gender == 1) / population.size
```

```
## [1] 0.1916817
```

```
sum(smoking == 1) / population.size
```

```
## [1] 0.5015069
```

```
p.ckd.if.female.smoking = 0.25
p.ckd.if.female.non.smoking = 0.10
p.ckd.if.male.smoking = 0.30
p.ckd.if.male.non.smoking = 0.15

ckd = vector("numeric", population.size)

for (k in 1:population.size) {
  if (gender[k] == 0 & smoking[k] == 1) {
    ckd[k] = sample(c(0, 1), prob = c(1 - p.ckd.if.male.smoking, p.ckd.if.male.smoking), size = 1)

  }

  if (gender[k] == 0 & smoking[k] == 0) {
    ckd[k] = sample(c(0, 1), prob = c(1 - p.ckd.if.male.non.smoking, p.ckd.if.male.non.smoking), size =

  }

  if (gender[k] == 1 & smoking[k] == 1) {
    ckd[k] = sample(c(0, 1), prob = c(1 - p.ckd.if.female.smoking, p.ckd.if.female.smoking), size = 1)

  }

  if (gender[k] == 1 & smoking[k] == 0) {
    ckd[k] = sample(c(0, 1), prob = c(1 - p.ckd.if.female.non.smoking, p.ckd.if.female.non.smoking), si
```

```
  }

}

addmargins(table(gender, smoking, ckd))
```

```
## , , ckd = 0
##
##       smoking
## gender    0    1  Sum
##    0    290  364  654
##    1    445  247  692
##    Sum  735  611 1346
##
## , , ckd = 1
##
##       smoking
## gender    0    1  Sum
##    0     49  150  199
##    1     43   71  114
##    Sum   92  221  313
##
## , , ckd = Sum
##
##       smoking
## gender    0    1  Sum
##    0    339  514  853
##    1    488  318  806
##    Sum  827  832 1659
```

```
sum(ckd == 1) / population.size
```

```
## [1] 0.1886679
```

```
sum(gender == 0 & ckd == 1) / sum(ckd == 1)
```

```
## [1] 0.6357827
```