

Conditional Probability

Akanksha Tippi

09-19-2024

Topics

- Definition of conditional probability
- Simulation

A bag contains 3 red and 3 white balls. Two balls are drawn from the bag, one at a time; the first ball is not replaced before the second ball is drawn.

1. What is the probability of drawing a white ball on the first pick and a red on the second?

```
#define parameters
balls = rep(c("R", "W"), c(5,2))
number.draws = 2
replicates = 10000

#create empty vector to store results
successes = vector("numeric", replicates)

#set the seed for a pseudo-random sample
set.seed(5011)

#simulate the draws
for(k in 1:replicates){

  draw = sample(balls, size = number.draws, replace = FALSE)

  if(draw[1] == "W" & draw[2] == "R"){
    successes[k] = 1
  }
}

#view the results
#successes
table(successes)
```

```
## successes
##      0      1
## 7590 2410
```

```
#define parameters
balls = rep(c("R", "W"), c(3,3))
number.draws = 2
replicates = 10000

#create empty vector to store results
successes = vector("numeric", replicates)

#set the seed for a pseudo-random sample
set.seed(2018)

#simulate the draws
for(k in 1:replicates){

  draw = sample(balls, size = number.draws, replace = FALSE)

  if(draw[1] == "W" & draw[2] == "R"){
    successes[k] = 1
  }

}

#view the results
table(successes)
```

```
## successes
##      0      1
## 6926 3074
```

```
#estimate the probability
sum(successes)/replicates
```

```
## [1] 0.3074
```

```
#define parameters
balls = rep(c("R", "W"), c(3,3))
number.draws = 2
replicates = 10000

#create empty vector to store results
successes = vector("numeric", replicates)
```

```

#set the seed for a pseudo-random sample
#set.seed(2018)

#simulate the draws

for(k in 1:replicates){

  draw = sample(balls, size = number.draws, replace = FALSE)

  if( (draw[1] == "W" & draw[2] == "R") | (draw[1] == "R" & draw[2] == "W") ){
    successes[k] = 1
  }
}

#view the results
table(successes)

```

```

## successes
##      0      1
## 3946 6054

```

```

#estimate the probability
sum(successes)/replicates

```

```

## [1] 0.6054

```

```

#define parameters
p.female = 0.50
p.tall.if.female = 0.03
p.tall.if.male = 0.20
population.size = 10000

#create empty vectors to store results
sex = vector("numeric", population.size)
tall = vector("numeric", population.size)

#set the seed for a pseudo-random sample
#set.seed(2018)

#assign sex
sex = sample(c(0,1), size = population.size, prob = c(1 - p.female, p.female),
             replace = TRUE)

#assign tall or not
for (k in 1:population.size){

```

```

if (sex[k] == 0) {
  tall[k] = sample(c(0,1), prob = c(1 - p.tall.if.male, p.tall.if.male),
                  size = 1, replace = TRUE)
}

if (sex[k] == 1) {
  tall[k] = sample(c(0,1), prob = c(1 - p.tall.if.female, p.tall.if.female),
                  size = 1, replace = TRUE)
}
}

#view results
addmargins(table(sex, tall))

```

```

##      tall
## sex      0      1  Sum
##  0    3974    990 4964
##  1    4906    130 5036
##  Sum    8880   1120 10000

```

```

#probability of female and tall
sum(tall == 1 & sex == 1)/population.size

```

```

## [1] 0.013

```

```

#probability of tall
sum(tall)/population.size

```

```

## [1] 0.112

```

```

#define parameters
p.disease.AA = 0.8
p.disease.Aa = 0.4
p.disease.aa = 0.1

p.AA = 0.81
p.Aa = 0.18
p.aa = 0.01

population.size = 10000

#create empty vectors to store results
genotype = vector("numeric", population.size)
disease = vector("numeric", population.size)

```

```

#set the seed for a pseudo-random sample
#set.seed(2018)

#assign genotype
genotype = sample(c("AA", "Aa", "aa"), size = population.size,
                  prob = c(p.AA, p.Aa, p.aa), replace = TRUE)

#assign disease status
for(k in 1:population.size){

  if(genotype[k] == "AA"){
    disease[k] = sample(c(0, 1), size = 1,
                       prob = c(1 - p.disease.AA, p.disease.AA),
                       replace = TRUE)
  }

  if(genotype[k] == "Aa"){
    disease[k] = sample(c(0, 1), size = 1,
                       prob = c(1 - p.disease.Aa, p.disease.Aa),
                       replace = TRUE)
  }

  if(genotype[k] == "aa"){
    disease[k] = sample(c(0, 1), size = 1,
                       prob = c(1 - p.disease.aa, p.disease.aa),
                       replace = TRUE)
  }
}

#view results
addmargins(table(genotype, disease))

```

```

##          disease
## genotype      0      1    Sum
##      aa      93      7    100
##      Aa    1105    689   1794
##      AA    1690   6416   8106
##      Sum    2888   7112  10000

```

```
sum(disease)/population.size
```

```
## [1] 0.7112
```

```
sum(genotype == "AA" & disease == 1)/sum(disease)
```

```
## [1] 0.9021372
```