

P3

Jeremiah Theisen

2024-09-14

```
smallpox = read.csv("https://github.com/TienChih/tbil-stats/raw/main/data/smallpox.csv")
names(smallpox)
```

```
## [1] "result"      "inoculated"
```

2.3.1

- a. $(C) = 70$
- b. $(S|C) = 56$
- c. $(!S|C) = 14$
- d. $(!C) = 30$
- e. $(S|!C) = 12$
- f. $(!S|!C) = 18$
- g. Venn Diagram
- h. $(S) = 68$
- i. $P(S) = 0.68$
- j. $P(C|S) = 0.824$
- k. $P(!C|!S) = 0.5625$

2.3.2

a.

```
cream = sample(c("cream", "no cream"), 100, replace = TRUE, prob = c(0.7, 0.3))
```

b.

```
sugar = rep(NA, 100)

for(i in 1:100){
  if(cream[i]=="cream"){
    sugar[i]=sample(c("sugar", "no sugar"), 1, replace = TRUE, prob = c(0.8, 0.2))
  }
  else{
    sugar[i]=sample(c("sugar", "no sugar"), 1, replace = TRUE, prob = c(0.4, 0.6))
  }
}

sugar
```

```
## [1] "no sugar" "sugar" "no sugar" "no sugar" "no sugar" "sugar"
## [7] "sugar" "no sugar" "sugar" "no sugar" "sugar" "no sugar"
## [13] "sugar" "no sugar" "sugar" "sugar" "sugar" "sugar"
## [19] "sugar" "sugar" "sugar" "sugar" "no sugar" "sugar"
## [25] "sugar" "no sugar" "no sugar" "sugar" "no sugar" "no sugar"
## [31] "sugar" "no sugar" "sugar" "no sugar" "sugar" "sugar"
## [37] "sugar" "sugar" "no sugar" "no sugar" "no sugar" "sugar"
## [43] "sugar" "sugar" "no sugar" "no sugar" "sugar" "sugar"
## [49] "sugar" "sugar" "no sugar" "sugar" "sugar" "sugar"
## [55] "sugar" "sugar" "sugar" "no sugar" "sugar" "sugar"
## [61] "no sugar" "sugar" "sugar" "sugar" "sugar" "sugar"
## [67] "no sugar" "no sugar" "sugar" "sugar" "sugar" "sugar"
## [73] "no sugar" "no sugar" "no sugar" "sugar" "sugar" "no sugar"
## [79] "no sugar" "sugar" "sugar" "sugar" "sugar" "sugar"
## [85] "no sugar" "no sugar" "sugar" "sugar" "no sugar" "sugar"
## [91] "sugar" "sugar" "no sugar" "sugar" "no sugar" "sugar"
## [97] "sugar" "no sugar" "sugar" "sugar"
```

c.

```
table(cream,sugar)
```

```
##          sugar
## cream    no sugar sugar
##  cream          13   47
## no cream         23   17
```

d. The values are off, but close enough given the sample size.

2.3.3

- $P(S) = 68$, this is the same value calculated earlier
- $P(S) = 0.48$
- $P(S) = 0.6$
- $P(S) = 0.4$
- $P(S) = 0.8$
- The visuals tell me that as $P(C)$ goes up, so does $P(S)$, as $P(S|C)$ is higher than $P(S|\bar{C})$. The values of X line up well.

2.3.4

- $P(Z|S)$
- $P(Z|\bar{S})$
- $P(Z)$
- SKIP
- SKIP

2.3.5

- $p(\text{innoculated}) = 0.0392$

```
length(which(smallpox$inoculated=="yes"))
```

```
## [1] 244
```

b. $P(N) = 0.96$

```
inocpatients=subset(smallpox, smallpox$inoculated=="yes")
notinocpatients=subset(smallpox, smallpox$inoculated=="no")
nrow(notinocpatients)
```

```
## [1] 5980
```

c. $P(D|I) = 0.024$

```
length(which(inocpatients$result=="died"))
```

```
## [1] 6
```

d. $P(D|N) = 0.1411$

```
length(which(notinocpatients$result=="died"))
```

```
## [1] 844
```

e. $P(I|D) = 6.898 * 10^{-3}, 0.0007$

f.

```
diedpatients=subset(smallpox, smallpox$result=="died")
nrow(diedpatients)
```

```
## [1] 850
```

g.

```
length(which(diedpatients$inoculated=="yes"))
```

```
## [1] 6
```

$P(I|D) = 7.06 * 10^{-3}; 0.0007$. This is the same as was calculated earlier when simplified. Given how many decimal places were involved with the earlier equation, I would say that this is close enough. I did the calculation for e twice and got a similar number both times.