Stats Theory 1: HW 1

Tim Vigers

8/30/2018

Problem 1

Using these tabulated data, find the exact numerical values for the following probabilities:

a. P(C|H & D): The number of Haitian IV drug users (H & D) is 150. Of these 100 are carriers, so

$$\frac{100}{150} = \frac{2}{3}$$

b. P(C or D|H): Since it's given that we're looking at Haitian men, you just find P(C or D) using the data for only Haitian men:

$$P(C \cup D) = P(C) + P(D) - P(C \cap D) = \frac{110}{200} + \frac{150}{200} - \frac{100}{200} = \frac{160}{200} = 0.8$$

c.

$$P(H|C^c) = \frac{P(H \cap C^c)}{P(C^c)} = \frac{50}{300} / \frac{140}{300} = \frac{5}{14}$$

d.

$$P((C \cap H)^c|D) = \frac{P((C \cap H)^c \cap D)}{P(D)} = \frac{110}{300} / \frac{210}{300} = \frac{11}{21}$$

e.

$$P(C \cup D \cup H) = P(C) + P(D) + P(H) - P(C \cap D) - P(D \cap H) - P(C \cap H) + P(C \cap D \cap H) =$$

$$\frac{160}{300} + \frac{210}{300} + \frac{200}{300} - \frac{140}{300} - \frac{150}{300} - \frac{110}{300} + \frac{100}{300} = \frac{270}{300} = \frac{9}{10}$$

This makes sense because the complement to this is everyone who is not Haitian, not a carrier, and not an IV drug user, of which there are only 30.

f.

$$P(C \cup (D \cap H)) = P(C) + P(D \cap H) - P(C \cap (D \cap H)) = \frac{160}{300} + \frac{150}{300} - \frac{100}{300} = \frac{210}{300}$$

Problem 2

a. This can be solved using the R binomial function qbinom(), which takes a probability value and returns the value cutoff for that quantile.

```
qbinom(p = 0.98,size = 10,prob = 0.75)
```

```
## [1] 10
```

b.

Learning objectives: dbinom, choose, and a for loop

If you need help with a function, type in help("function")

2b

A particular cancer experiment requires at least two highly susceptible mice. What is the probability that an order of 10 mice from the breeding facility will suffice?

Binomial formula: p(x) = choose(n, x) p^x (1-p)^(n-x)

Binomial R function: dbinom(x, size=n, prob, log = FALSE)

```
1 - dbinom(size = 10, prob = 0.75, x = 0) -  dbinom(size = 10, prob = 0.75, x = 1)
```

```
## [1] 0.9999704
```

2c

Suppose that a particular batch of 7 mice actually contains exactly 3 highly susceptible ones. If the cancer researcher chooses three mice at random from this batch for a particular experiment, what is the probability that at least two of the three mice chosen will be highly susceptible ones?

choose function in R: choose(n, k) with n as size and k an integer

```
1 - (choose(3,0) * choose(4,3)) / choose(7,3) - (choose(3,1) * choose(4,2)) / choose(7,3)
```

```
## [1] 0.3714286
```

```
13/35
```

```
## [1] 0.3714286
```

2d

Notice that we can partition the event space into four possible subsets:

i) 0 out of the 3 mice are highly susceptible.

```
dbinom(size = 3, prob = 0.05, x = 2)
```

```
## [1] 0.007125
```

ii) 1 out of the 3 mice is highly susceptible.

```
dbinom(size = 2, prob = 0.05, x = 1) * 0.15 + (0.05)^2 * 0.85
```

```
## [1] 0.016375
```

iii) 2 out of the 3 mice are highly susceptible.

```r
choose(2,1) * (0.15) * (0.05) * (0.85) + (0.15)^2 * (0.95)
```

```
## [1] 0.034125
```

iv) 3 out of the 3 mice are highly susceptible.

```r
choose(3,2) * (0.15)^2 * (0.85)
```

```
## [1] 0.057375
```

For loop in R: for(var in seq) expr

for some variable var in the sequence seq, do the expression expr

```r
for(i in 0:3) print((choose(3,i) * choose(4, 3-i)) / choose(7,3))
```

```
## [1] 0.1142857
## [1] 0.5142857
## [1] 0.3428571
## [1] 0.02857143
```

```r
4/35
```

```
## [1] 0.1142857
```

```r
18/35
```

```
## [1] 0.5142857
```

```r
12/35
```

```
## [1] 0.3428571
```

```r
1/35
```

```
## [1] 0.02857143
```