Homework 05 - Probability

Problem 1

Suppose you roll three fair dice (one at a time - so you know the order in which each number appears).

Part a)

How many different outcomes are possible?

16 different sums, 6x6x6 = 216 permutations of the three dice.

Part b)

What values are possible for the sum on the three dice?

$\{x \mid 3 \leq x \leq 18,\, x \in \mathbb{Z}\}$

Part c)

Give the number of outcomes that result in a sum of 2, 3, 4, and 5?

2: 0 outcomes

3: 1 outcome ({1,1,1})

4: 3 outcomes ({1,1,2}, {1,2,1}, {2,1,1})

5: 6 outcomes ({1,2,2}, {2,2,1}, {2,1,2}, {1,1,3}, {1,3,1}, {3,1,1})

Part d)

Find $P\{\text{sum is } x\}$ for $x = 2, 3, 4$, and 5.

P(sum is 2) = 0/216

P(sum is 3) = 1/216

P(sum is 4) = 3/216

P(sum is 5) = 6/216

Part e)

Find $P\{\text{sum is even}\}$. *Hint: Consider the cases for which the values on each of the three die are even or odd.*

50% of the sums are even because each dice has a 50% chance of odd or even.

Part f)

Use the sample command with the parameter `replace=TRUE` to simulate 10000 times the sum on the three dice. Give a table of outcomes and compare the proportion in the simulation to the probabilities computed in part c).

```
# We create the probabilities based on the number of possible outcomes for each values
three_dice_probs <- c(1, 3, 6, 10, 15, 21, 25, 27, 27, 25, 21, 15, 10, 6, 3, 1) / 216

#sample(values to sample, number of samples, replace (yes or no), probability vector)
samp_1 <- sample(3:18, 10000, replace = TRUE, prob = three_dice_probs)

#table will create a table of values for each of the possible options
prop.table(table(samp_1))
```

```
samp_1
     3      4      5      6      7      8      9     10     11     12     13 
0.0040 0.0118 0.0286 0.0452 0.0700 0.0964 0.1106 0.1243 0.1255 0.1192 0.0997 
    14     15     16     17     18 
0.0719 0.0475 0.0276 0.0119 0.0058 
```

Part g)

Explain, in words, why $P\{\text{sum is } x\} = P\{\text{sum is } 21 - x\}$. *Hint: If the value v is showing on the die, then the value $7 - v$ is opposite.*

Because you can just flip the die over. If v is showing on the dice, the dice flipped over is 7-v. 1 is showing, flip, then 6. 7-1 = 6

Problem 2

With the monsoon season, we can have more cases of dengue fever, which is a mosquito-borne tropical disease caused by the dengue virus. Antibody tests are recommended during a dengue outbreak. However, the presence of other viruses in the human body can have cross-reactive results, yielding a high false positive rate.

- Assume a false positive rate of 10% (i.e., $P\{\text{Test+} \mid \text{no dengue}\} = 0.10$)
- Assume a false negative rate of 1% (i.e., $P\{\text{Test-} \mid \text{has dengue}\} = 0.01$))

Part a)

Given that a person has dengue, what is the probability of a positive test? 100% - false negative = 99%

Part b)

If two percent of a population has dengue (i.e., $P\{\text{has dengue}\} = 0.02$), what fraction of the population will test positive? 0.02 * 0.99 + 0.1 * 0.98 = 0.1178

Part c)

If the individual tests positive, what is the probability that this individual has dengue? In other words, determine $P\{\text{has dengue} \mid \text{Test+}\}$. Assume the same population dengue rate as in part (b) (i.e., $P\{\text{has dengue}\} = 0.02$). 0.99 * 0.02 / 0.1178 = 0.1680814941

Part d)

Produce a two-way table showing the numbers of people with and without dengue and with and without a positive test out of a population of 10,000. Round each entry to the nearest integer. Assume the same population dengue rate as in part (b) (i.e., $P\{\text{has dengue}\} = 0.02$).

```r
# Population
n <- 10000

# Given probabilities
p_dengue <- 0.02
p_no_dengue <- 1 - p_dengue
p_pos_given_dengue <- 0.99
```

```r
p_pos_given_no_dengue <- 0.10

# Cell counts
has_dengue <- n * p_dengue          # 200
no_dengue  <- n * p_no_dengue       # 9800

tp <- round(has_dengue * p_pos_given_dengue)    # True positive
fn <- round(has_dengue * (1 - p_pos_given_dengue)) # False negative
fp <- round(no_dengue  * p_pos_given_no_dengue)  # False positive
tn <- round(no_dengue  * (1 - p_pos_given_no_dengue)) # True negative

# Build the table
two_way <- matrix(c(tp, fp, tp + fp,
                    fn, tn, fn + tn,
                    tp + fn, fp + tn, n),
                  nrow = 3, ncol = 3,
                  dimnames = list(
                    c("Test+", "Test-", "Total"),
                    c("Has Dengue", "No Dengue", "Total")
                  ))

two_way
```

```
      Has Dengue No Dengue Total
Test+        198         2   200
Test-        980      8820  9800
Total       1178      8822 10000
```

Part e)

Confirm your answer in part (c) by calculating the corresponding proportion using the information from the two way table in part (d). Confirmed - 198/1178 = 0.1680814941