

A_M1

October 18, 2024

0.0.1 Problem 1:

Suppose we are interested in the buying habits of shoppers at a particular grocery store with regards to whether they purchase apples, milk, and/or bread. Now suppose 30% of all shoppers at this particular grocery store buy apples, 45% buy milk, and 40% buy a loaf of bread. Let A be the event that a randomly selected shopper buys apples, B be the event that the same randomly selected shopper buys milk, and C the event that they buy bread. Suppose we also know (from data collected) the following information:

- The probability that the shopper buys apples and milk is 0.20.
- The probability that the shopper buys milk and bread is 0.25.
- The probability that the shopper buys apples and bread is 0.12.
- The probability that the shopper buys all three items is 0.07.

Use this information to answer the following questions.

1 Module 1 Assessment

In this notebook you will apply the axioms of probability and use descriptive statistics to solve various problems. Please read each statement and answer all questions in the code blocks that contain “your code here”. Feel free to add as many blocks as you need, only the original code blocks will be graded.

You will see some cells contain the comment “Hidden Test Cell.” These cells contain code to test your solution, which will run after you submit your assignment. You will not know what tests are being run, or if they have been passed until you have submitted your assignment.

```
[1]: #Imports test that library
library(testthat)
```

a) For our purposes, we will use a numeric representation for each event. For example, (010) would be an event in the sample space where a zero in the first place represents no apples were bought while a 1 means they were. Similarly, the second place is the presence of milk and the third place of bread. The example given (010) represents the purchase of milk but not apples or bread.

Insert into vector S the events that belong to the sample space. Then insert the events from the sample space that would correspond to A occurring into vector A . Repeat this with vector B for B and vector C for C .

According to set notation, the sample space should look like this: $S = \{\dots\}$. However, due to data storage syntax in R, we will be storing these events in vectors. For example, for some arbitrary event W would be stored as follows $W = c(010)$. `c()` is a command we can use to construct a vector where commas separate each entry. Complete the code below. Do not worry about the order in which events are placed in vector.

```
[2]: # Sample Space
S <- c(000, 001, 010, 011, 100, 101, 110, 111)

# Event A: Buys apples (1 in the first position)
A <- c(100, 101, 110, 111)

# Event B: Buys milk (1 in the second position)
B <- c(010, 011, 110, 111)

# Event C: Buys bread (1 in the third position)
C <- c(001, 011, 101, 111)
```

```
[3]: # Test cell
```

```
[4]: # Hidden test cell
```

```
[5]: # Hidden test cell
```

```
[6]: # Hidden test cell
```

b) Find the probability that the shopper purchases at least one of the three items.

Store your answer in a variable `p1.1`.

```
[7]: p1.1 <- 0.65
```

0.65

```
[8]: # Hidden test cell
```

c) Find the probability that the shopper purchases none of the three items.

Store your answer in a variable `p1.2`.

```
[9]: p1.2 <- 0.35
```

0.35

```
[10]: # Hidden test cell
```

d) Find the probability that the shopper buys milk and bread but not apples.

Store your answer in a variable `p1.3`.

```
[11]: p1.3 <- 0.18
```

0.18

```
[12]: # Hidden test cell
```

1.0.1 Problem 2:

A student takes a multiple choice test with 20 questions. Each question has 5 possible answers, only one of which is correct.

a) How many ways can the test be completed? (Find the cardinality of the sample space.) Store your answer in `cardinality`.

```
[13]: cardinality <- 5^20
```

95367431640625

```
[14]: # Hidden test cell
```

b) If a student answers each question at random, what is the probability that they will answer at least 14 questions correctly? Round your answer to seven decimal places. Store your answer in `p2.b`.

```
[15]: p2.b <- sum(dbinom(14:20, size = 20, prob = 0.2))  
p2.b <- round(p2.b, 7)
```

1.84500572127232e-06

```
[16]: # Hidden test cell
```

c) If a student knows the answer to each question with probability 0.9, what is the chance they will answer at least 14 correctly? Round your answer to four decimal places. Save your answer as `p2.c`.

```
[17]: p2.c <- sum(dbinom(14:20, size = 20, prob = 0.9))  
p2.c <- round(p2.c, 4)
```

0.997613910591034

```
[18]: # Hidden test cell
```

Thought question (Ungraded): If a student understands 90% of the material, is a multiple choice test of 20 questions an appropriate means of testing?

1.0.2 Problem 3:

Suppose there are 20 employees on the day shift, 15 on swing, and 10 on nights. You would like to choose 6 at random for an in-depth interview. Let A_1 be the event that all 6 are chosen from the day shift, A_2 be the event that all 6 are chosen from the swing shift, and A_3 the event that all 6 are chosen from the night shift.

a) What is the cardinality of the sample space (the drawing of 6 employees from all shifts)? Assume each employee is unique. Think about what each element, or possible outcome, in the sample space represents.

```
[19]: Size <- choose(45, 6)
```

8145060

```
[20]: # Hidden test cell
```

b) Assuming that each employee is chosen with equal probability, what is the probability that all 6 chosen employees work the day shift? Note that the employees are chosen sequentially, not all at once. Round your answer to 5 decimal places. Save your answer as P_AllDay.

Thought question (ungraded): If you hired a consultant to do this interview and all 6 employees were chosen from the day shift, what would you be likely to conclude?

```
[21]: P_AllDay <- choose(20, 6) / choose(45, 6)
      P_AllDay <- round(P_AllDay, 5)
```

0.00475871264300079

```
[22]: # Hidden test cell
```

c) What is the probability that all 6 workers are selected from the same shift? Again, assume the employees are selected sequentially and not all at once. Round your answer to four decimal places. Save your answer as P_AllSame.

```
[23]: P_AllSame <- (choose(20, 6) + choose(15, 6) + choose(10, 6)) / choose(45, 6)
      P_AllSame <- round(P_AllSame, 4)
```

0.00539897803085551

```
[24]: # Hidden test cell
```

Thought question (Ungraded): Let B_1 be the event that no one from the day shift is selected, B_2 the event that no one from the swing shift is selected, and B_3 the event that no one from the night shift is selected.

How is A_1^c defined? Is it related to B_1 or B_1^c ? It's very important to carefully define the events of interest so that you know you are calculating the correct probability.

d) Find the probability that at least one of the shifts will be unrepresented in the sample of 6 workers. Round your answer to four decimal places. Save your answer as P_NoneOfOneShift.

Hint: We want to solve $P(B_1 \cup B_2 \cup B_3)$. To solve this, it may help to make a venn diagram for these events. Make sure you don't overcount the overlapping areas of this venn diagram!

```
[25]: P_Noneofoneshift <- 1 - (1 - P_AllDay - (choose(15, 6) / choose(45, 6)) -
  ↪ (choose(10, 6) / choose(45, 6)) + (choose(35, 6) / choose(45, 6)) +
  ↪ (choose(30, 6) / choose(45, 6)) + (choose(25, 6) / choose(45, 6)) -
  ↪ (choose(20, 6) / choose(45, 6)))
```

```
P_Noneofoneshift <- round(P_Noneofoneshift, 4)
```

0.994601021969144

```
[26]: # Hidden test cell
```