P_M3_1

October 25, 2021

1 Module 3 Peer Review Assignment

2 Problem 1

You work at a factory that manufactures light bulbs. You have determined that 5% of light bulbs that are produced are defective. For each of the scenarios below:

1. Define an appropriate random variable and distribution.
2. State the values that the random variable can take on.
3. State any assumptions that you need to make.
4. Find the probability that the random variable you defined takes on the value $X = 4$.

Part a)

Out of 30 lightbulbs, k are defective.

1. Let X be the number of lightbulbs that are defective. The distributions is $X \sim Bin(30, 0.05)$
2. Letting n represent the total number of lightbulbs, the values X can take on are {0, ..., n}
3. we need to make sure that each trial is independent. In other words, the probability of one light bulb being defective doesn't affect another.
4. As given by the code below, the probability that 4 of 30 lightbulbs is defective is 0.04514

[2]:
```
dbinom(4, 30, 0.05)
```

0.0451360511802104

Part b)

You test each lightbulb as it comes of the line. The k^{th} light bulb is defective.

1. Let X be the number of lightbulbs that are defective. The distributions is $X \sim FS(0.05)$
2. Letting n represent the total number of lightbulbs, the values X can take on are {1, ..., n}
3. We need to make sure that each trial is independent. In other words, the probability of one light bulb being defective doesn't affect another. Also, we need to make sure the probability of success is the same for each trial.
4. As given by the code below, the probability the 4^{th} lightbulb is the first defective one is .0429.

[5]:
```
# For first success, use the geomertic function but make x one less since the␣
↪geometric distribution
# doesn't include the sucessful trial
```

```
dgeom(3, 0.05)
```

0.04286875

Part c)

You find your second defective drive after observing k drives in all.

1. Let X be the number of lightbulbs that are defective. The distributions is $X \sim NBin(2, 0.05)$
2. Letting n represent the total number of lightbulbs, the values X can take on are {2, …, n}.
3. We need to make sure that each trial is independent. In other words, the probability of one light bulb being defective doesn't affect another.
4. As given by the code below, the probability we find our second defective driver after observing 4 drives in all is 0.0068

[8]:
```
# We will use the negative binomial distribution
# dbinom(number of failures, number of successes we want, probability)
dnbinom(2, 2, .05)
```

0.00676875

3 Problem 2

Consider a loaded six-sided die that is twice as likely to roll an even number as an odd number. Let X be random variable for value that is rolled from the die.

Part a)

What is the Probability Mass Function for X. Write this out as a table.

The PMF for the faces on the dies is shown below

[15]:
```
pmf = matrix(c((1/9), (2/9), (1/9), (2/9), (1/9), (2/9)), ncol=6, byrow = TRUE)
colnames(pmf) = c('1', '2', '3', '4', '5', '6')
pmf = as.table(pmf)
pmf
```

```
          1         2         3         4         5         6
A 0.1111111 0.2222222 0.1111111 0.2222222 0.1111111 0.2222222
```

Part b)

What is the Cumulative Distribution Function for X?

Let F(y) represent the CDF for that value y

$$F(y) = \begin{cases} 0 & \text{if y} < 1 \\ \frac{1}{9} & \text{if } 1 \leq \text{y} < 2 \\ \frac{3}{9} & \text{if } 2 \leq \text{y} < 3 \\ \frac{4}{9} & \text{if } 3 \leq \text{y} < 4 \\ \frac{6}{9} & \text{if } 4 \leq \text{y} < 5 \\ \frac{7}{9} & \text{if } 5 \leq \text{y} < 6 \\ 1 & \text{if } 6 \leq \text{y} \end{cases} \tag{1}$$

Part c)

What is $E[X]$?

E[X] = 3

[17]:
```
(1/9)*1 + (2/9)*2 + (1/9)*3 + (2/9)*4 + (1/9)*5 + (1/9)*6
```

3

4 Problem 3

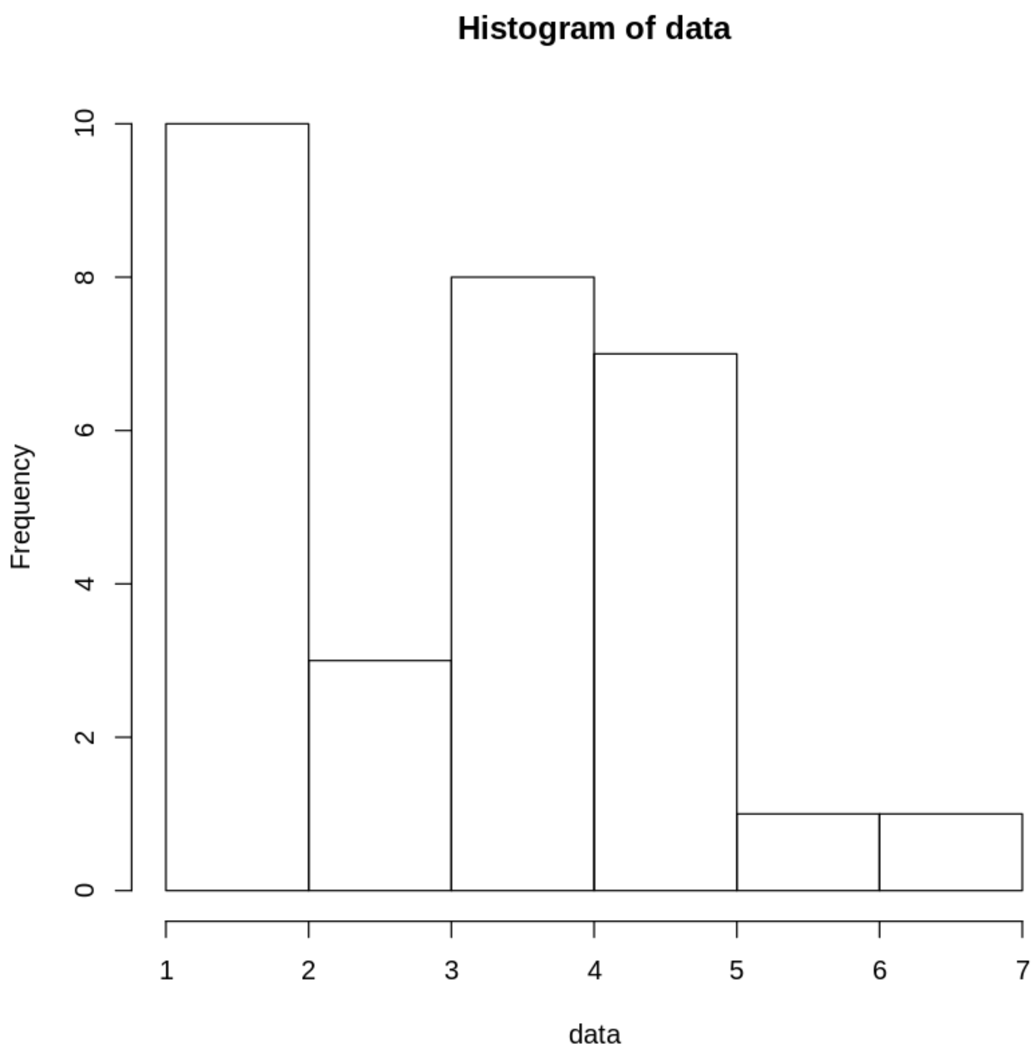
How would we simulate variables from these distributions in R? It'll turn out that the method is fairly similar across all these distributions so, for simplicity, let's just say we want to simulate $X \sim Bin(n, p)$. Take a look at the official documentation for this function here. Not extremely clear, is it? Let's go through it one step at a time.

Part a)

What if we want a random variable from this distribution? That is, we know some underlying distribution and we want to simulate many results from that distribution. Then we would use the "random generation" function `rbinom()`.

Play around with this function, with different `size` and `prob` parameters to get a feel for how it works. Finally, generate 30 results from a $Bin(10, 0.3)$ distribution and plot a histogram of the results.

[26]:
```
set.seed(23744)
data = rbinom(30, 10, 0.3)
hist(data)
```



Part b)

What if we have some value k and we want to know what's probability of generating k? That is, we're solving the Probability Mass Function $P(X = k)$. Then we would use the "density" function `dbinom()`.

Let $X \sim Bin(15, 0.4)$. By hand, solve $P(X = 4)$. Then use the `dbinom()` function to confirm your result.

Given $X \sim Bin(15, 0.4)$, $P(X = 4) = 0.1268$

[27]:
```
dbinom(4, 15, 0.4)
```

0.12677580324864

Part c)

What if we wanted to solve for some value of the Cumulative Density Function? That is, we know k and want to find $P(X \leq k) = p$. Then we would use the "distribution function" `pbinom()`.

Let $X \sim Bin(15, 0.4)$. By hand, solve $P(X \leq 4)$. Then use the `qbinom()` function to confirm your result.

Given $X \sim Bin(15, 0.4)$, $P(X \leq 4) = 0.2173$

[42]:
```
pbinom(4, 15, 0.4)
```

0.217277705650176

Part d)

Finally, we have the "quantile" function `qbinom()`. This function is the reverse of the `pbinom()` function, in that it takes a probability p as an argument and returns the value k of the CDF that results in that much probability.

Use the `qbinom()` function to confirm your results from **Part c**. That is, plug in the probablity you got from **Part c** and see if you get the same k.

[43]:
```
qbinom(0.212777, size=15, prob=.4)
```

4

Nearly every distribution has these four functions, and they will be very useful for our future calculations and simulations.