Homework 5

2024-03-26

Question 1

a)

We know that 3 out of every 75 lightbulbs are found to be defective, hence the probability of finding a defective lightbulb is $\frac{1}{25}$. This also means that the probability of finding a non-defective lightbulb is $\frac{24}{25}$.

So, the probability of Max finding a defective lightbulb on the 6th one he tests is:

```r
round((24/25)^5 * 1/25, 3)
```

```
## [1] 0.033
```

b)

The probability of taking at least 4 trials to find a defective lightbulb can be labelled as P($defective \geq 4$). This can also be calculated as 1 - P($defective \leq 4$).

1 - P($defective \leq 4$) = 1 - (P(1) + P(2) + P(3)), where P(1), P(2), and P(3), are the probabilities of finding a defective lightbulb in the 1st, 2nd and 3rd trials specifically.

Now, using part a), we can take the formula for the probability of finding a faulty lightbulb at a certain trial as:

P(n) = q^{n-1} x p, where q = $\frac{24}{25}$ and p = $\frac{1}{25}$

Also, P(1) = p.

So, probability of taking at least 4 trials is

```r
round(1 - (1/25) - (1/25 * 24/25) - (1/25 * (24/25)^2), 3)
```

```
## [1] 0.885
```

c)

Using our formula above,

The probability of taking at most 10 trials to find the first defective lightbulb is:

P($defective \leq 4$) = P(1) + P(2) + ... P(10)

```r
round(1 - (1/25) - (1/25 * 24/25) - (1/25 * (24/25)^2) -
        (1/25 * (24/25)^3) - (1/25 * (24/25)^4) - (1/25 * (24/25)^5) -
        (1/25 * (24/25)^6) - (1/25 * (24/25)^7) - (1/25 * (24/25)^8) -
        (1/25 * (24/25)^9), 3)
```

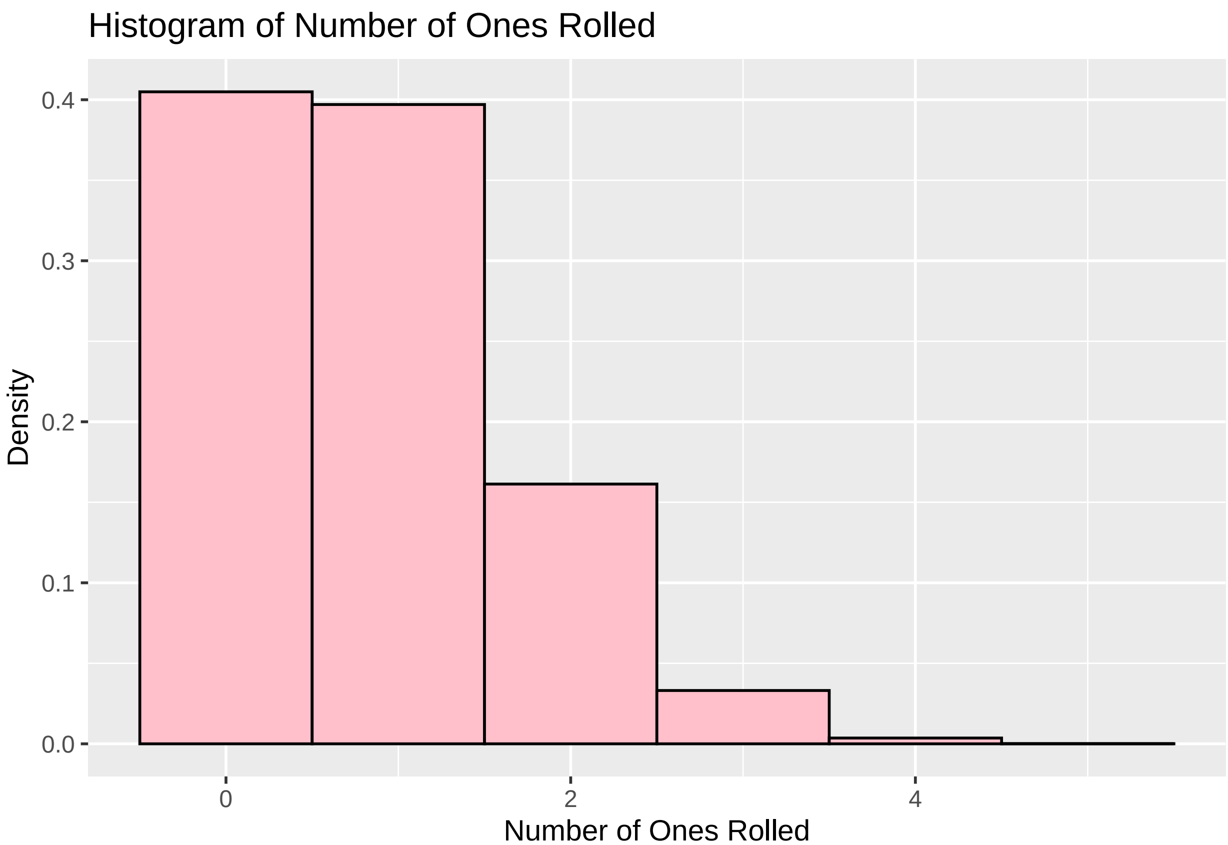
```
## [1] 0.665
```

Question 2

```r
set.seed(20220707)
X <- replicate(10000, sum(sample(1:6, 5, replace = TRUE) == 1))
```

```r
library(ggplot2)
ggplot(data.frame(X), aes(x = X)) +
  geom_histogram(binwidth = 1, fill = "pink", color = "black", aes(y = ..density..)) +
  labs(title = "Histogram of Number of Ones Rolled",
       x = "Number of Ones Rolled",
       y = "Density")
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```r
round(mean(X), 3)
```

```
## [1] 0.834
```

```r
round(var(X), 3)
```

```
## [1] 0.705
```

Question 3

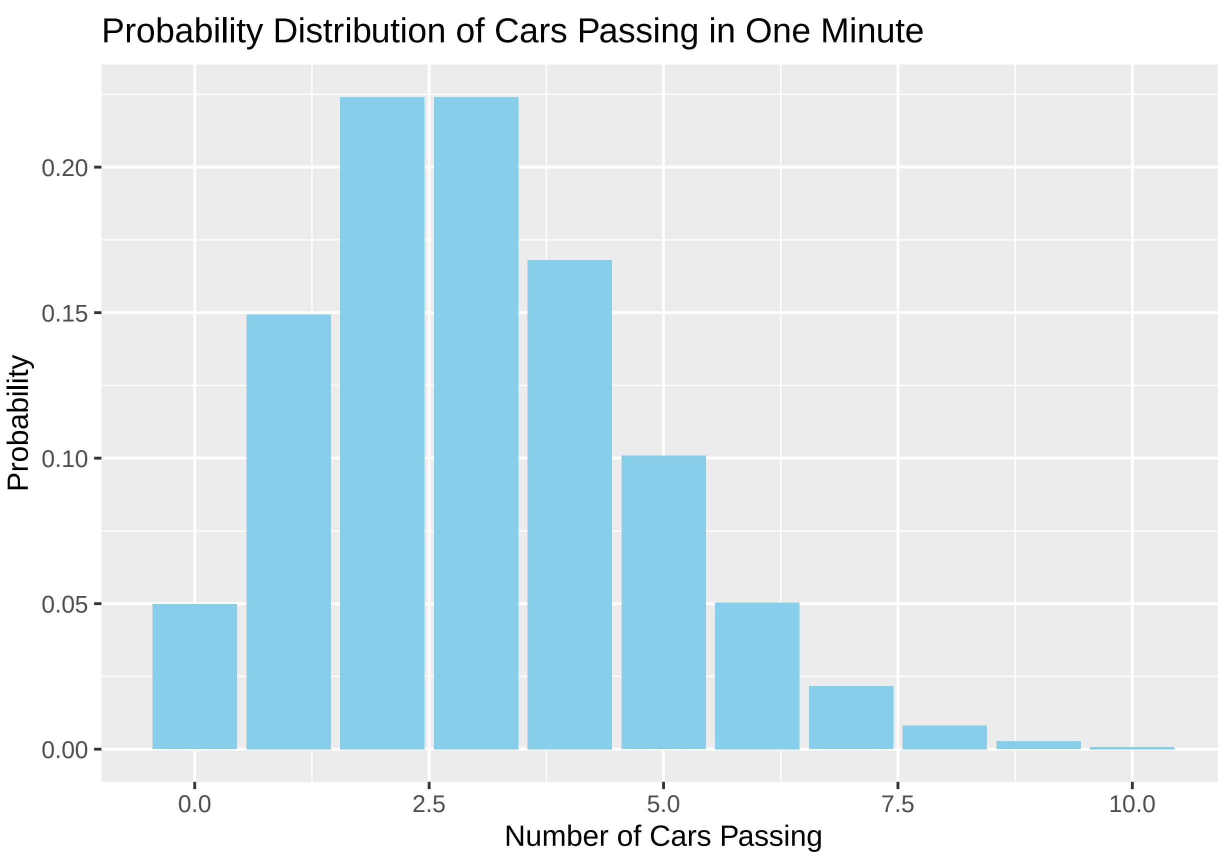
```r
# Calculating the probability of congestion (more than 5 cars in one minute)
lambda <- 180/60
prob_congestion <- round(1 - sum(dpois(0:5, lambda)), 3)
prob_congestion
```

```
## [1] 0.084
```

```r
library(ggplot2)

cars <- 0:10
prob <- dpois(cars, lambda)
data <- data.frame(cars, prob)

ggplot(data, aes(x = cars, y = prob)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(x = "Number of Cars Passing", y = "Probability") +
  ggtitle("Probability Distribution of Cars Passing in One Minute")
```



Question 4

```r
mean_score <- 500
std_dev <- 100

prob_585 <- round(pnorm(585, mean_score, std_dev), 3)
prob_585
```

```
## [1] 0.802
```

```r
lower_quartile <- round(qnorm(0.25, mean_score, std_dev), 3)
median_score <- round(qnorm(0.5, mean_score, std_dev), 3)
upper_quartile <- round(qnorm(0.75, mean_score, std_dev), 3)

lower_quartile
```

```
## [1] 432.551
```

```r
median_score
```

```
## [1] 500
```

```r
upper_quartile
```

```
## [1] 567.449
```

Question 5

Using Baye's Theorem,

The cabs have a 0.95 chance of being green, and a 0.005 chance of being red. Now, if the cab was green, the witness has a 0.2 chance of saying that it is red, and if the cab was red, the witness has a 0.8 chance of saying that the car was red.

We are looking to find how likely it is that the cab was red, given that the witness said it was red.

Our formula will be:

P(Red | Identified as Red) = $\frac{P(Red)*P(IdentifiedAsRed|Red)}{P(IdentifiedAsRed)}$

P(Identified as Red) = P(Identified as Red) * P(Red) + P(Identified as Red) * P(Green) = (0.8 * 0.05) + (0.2 * 0.95) = 0.23

So, P(Red | Identified as Red) = 0.05 * 0.8/ 0.23

```r
round((0.05 * 0.8)/0.23, 3)
```

```
## [1] 0.174
```