OREGON STATE UNIVERSITY

ST 314

SUMMER 2019

Data Analysis One

Author:
Thomas Noelcke

Instructor:
Katie Jager

I. Part 1

A. Random Variable 1

This variable would be best modeled by the exponential distribution. This is because it has no upper limit and is positively skewed. Additionally, we know the average but not the standard distribution. To best model this distribution we only need one parameter, the average. Given the average we can calculate the value for lambda. Show below is the probability density function.

$$f_x = \begin{cases} \frac{1}{100} e^{-\frac{1}{100}x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

B. Random Variable 2

This random variable would be best modeled by the normal distribution. The probability is symmetrical where it is more likely that the values are closer to the mean. The distribution of the dairy cows also has a calculated standard deviation which makes it easy to model with the normal distribution. For this distribution we need the average value for the distribution in question and the standard distribution. In this case the average is 4.5% and the standard deviation is 0.4%. Shown below is the probability density function.

$$f_x = \frac{1}{\sqrt{2\pi 0.4^2}} e^{-\frac{(x-4.5)^2}{0.8^2}}$$

C. Random Variable 3

This random variable would be best modeled by the Uniform distribution. This is because between 2pm and 4pm the pump was in a failure state. This distribution also works well because there are clear upper and lower bounds to the distribution and a single value during that period of time. In this cause we only need the upper and lower bounds for the distribution and the value for the distribution. In this case the upper and lower bounds are 2 to 4. We will let the chance of failure at any given one point during the the two hours be represented by the value 1/120. Below is the the probability density function.

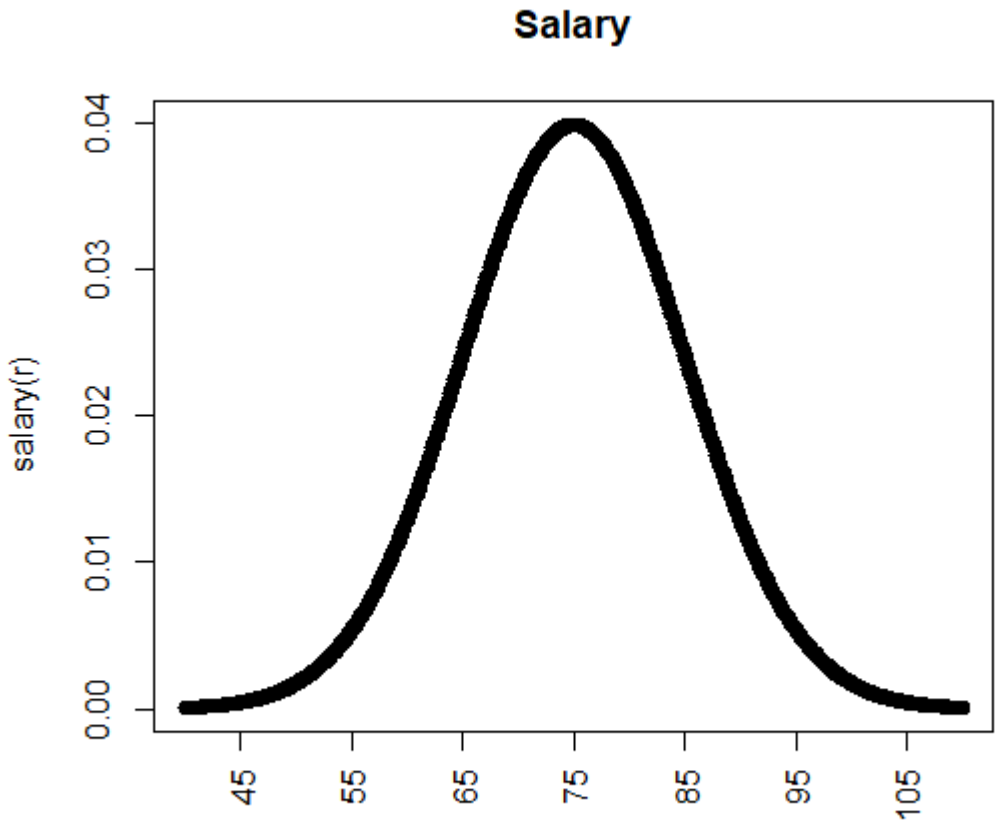
$$f_x = \begin{cases} \frac{1}{120} & 2 \geq x \leq 4 \\ 0 & x > 2, x < 4 \end{cases}$$

II. Part 2

A.

The average appears to be 75,000$ according to the distribution shown in the picture. based on the 65, 95 99 rule I would estimate that the standard deviation appears to be x. I've placed some R code and a recreation of the plot which I've used for the rest of the calculations below.

```
salary = function(x) {dnorm(x, mean = 75, sd=10)+75}
r = seq(40, 110, 0.1)
lab = seq(45, 105, 10)
plot(r, salary(r), xlab="", main="Salary", lwd = 2, xaxt="n")
axis(1, at = lab, las=2)
```



B.

For this problem we can simply calculate a Z Score and determine how far away that 59,000$ is away from the mean.

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{59000 - 75000}{10000} = -1.6$$

The salary of 59,000$ per year is 1.6 standard deviations less than the mean.

C.

To find the probability that an engineers salary is less than 59000 I used R. I've included the R code that I used to find the answer below as well as the output to the console.

```
salaryProp = function(x) {pnorm(x, mean = 75, sd = 10)}
salaryProp(59)
```

```
> salaryProp = function(x) {pnorm(x, mean = 75, sd = 10)}
> salaryProp(59)
[1] 0.05479929
```

D.

To find the 85th percentile of mechanical engineers salary I decided to user R. I've include the R code that i used to find the answer below as well as the output from the console.

```
salary = function(x) {qnorm(x, mean = 75, sd=10)}
salary(0.85)

> salary = function(x) {qnorm(x, mean = 75, sd=10)}
> salary(0.85)
[1] 85.36433
```

This means that the salary that is at the 85th precentile is 85,363.33$.

III. Part 3

A.

a) i: The expression of the probability density function is shown below:

$$f_x = \begin{cases} \frac{1}{7^{\alpha}\gamma(2)} x e^{\frac{-x}{7}} & x \geq 0 \\ 0 & otherwise \end{cases}$$

b) ii: The average and standard deviation are shown below. Average Expected Value:

$$E(x) = u_x = \alpha$$

$$E(x) = 14$$

Standard Deviation:

$$V(x) = \sigma_x^2 = \alpha^2$$

$$sd = sqrt2 * 7^2)$$

$$sd = 9.8994$$

c) iii: I chose to use R for this problem below is the r code I used as well as the results from the terminal.

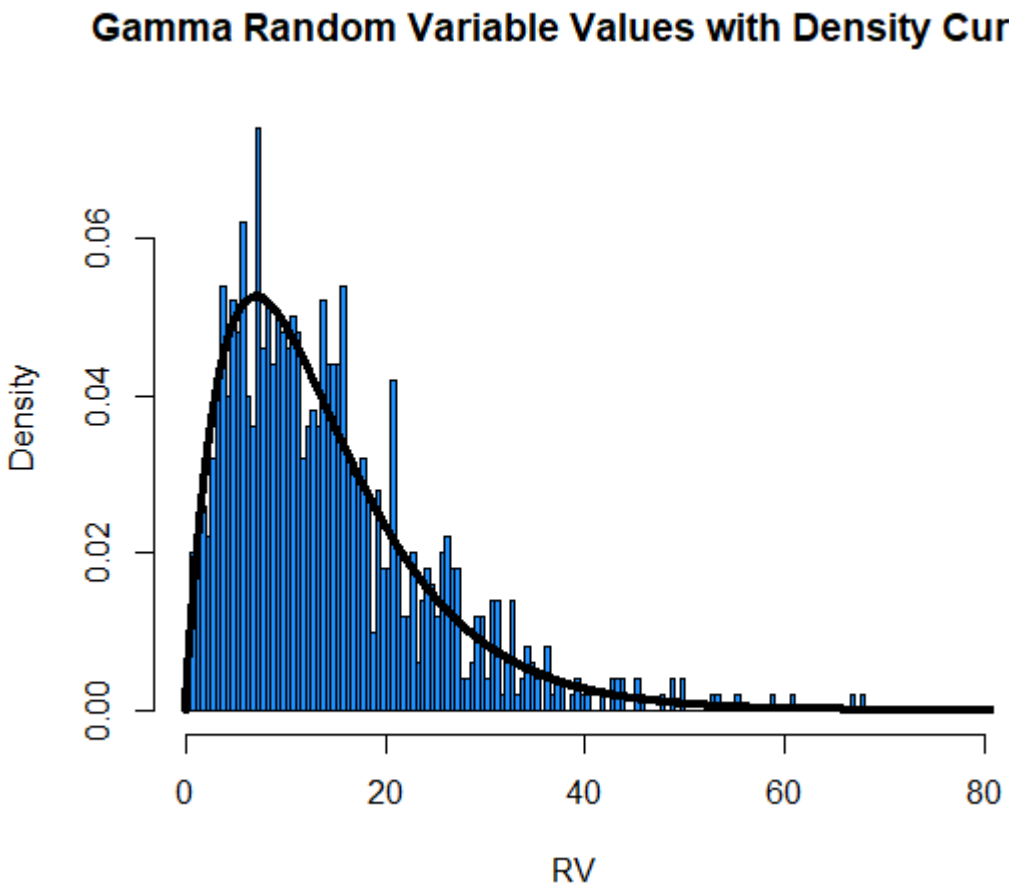
```
gamma = function(x) {pgamma(x,2, (1/7))}
gamma(4)

> gamma = function(x) {pgamma(x,2, (1/7))}
> gamma(4)
[1] 0.1125858
```

The probability that x is less than 4 is 0.1126.

B.

I've run the simulation and found that the simulation generally follows the shape of the curve. However, it has area's where it is way out side the curve and areas where it is way under the curve. Below is a picture of the plot my simulation generated.

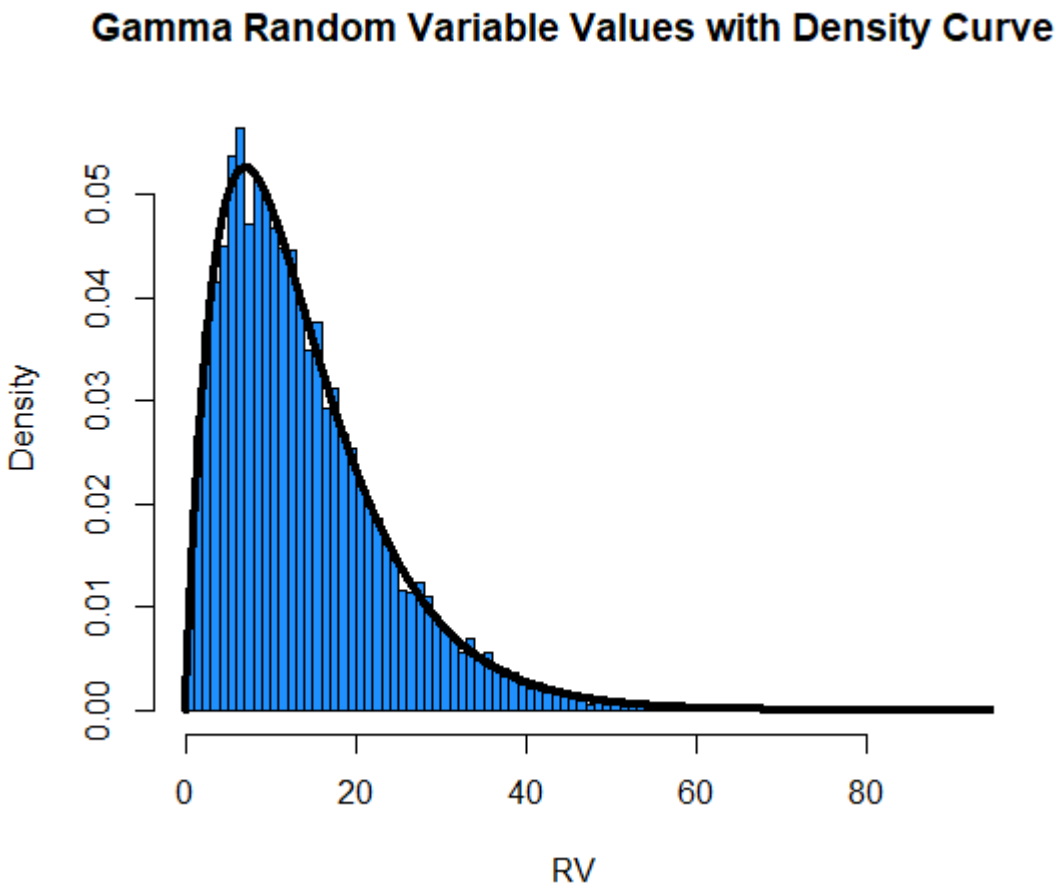


C.

The exact proportion below 4 for the simulation was 0.11. This matches fairly closely with the expected value found in part 2.iii.

D.

The increasing number of trials has had the effect of a much tighter fit to the curve. There are still some values that are inside or outside the curve. However, Those values are much less extreme than the trial where we only ran one thousand trials.

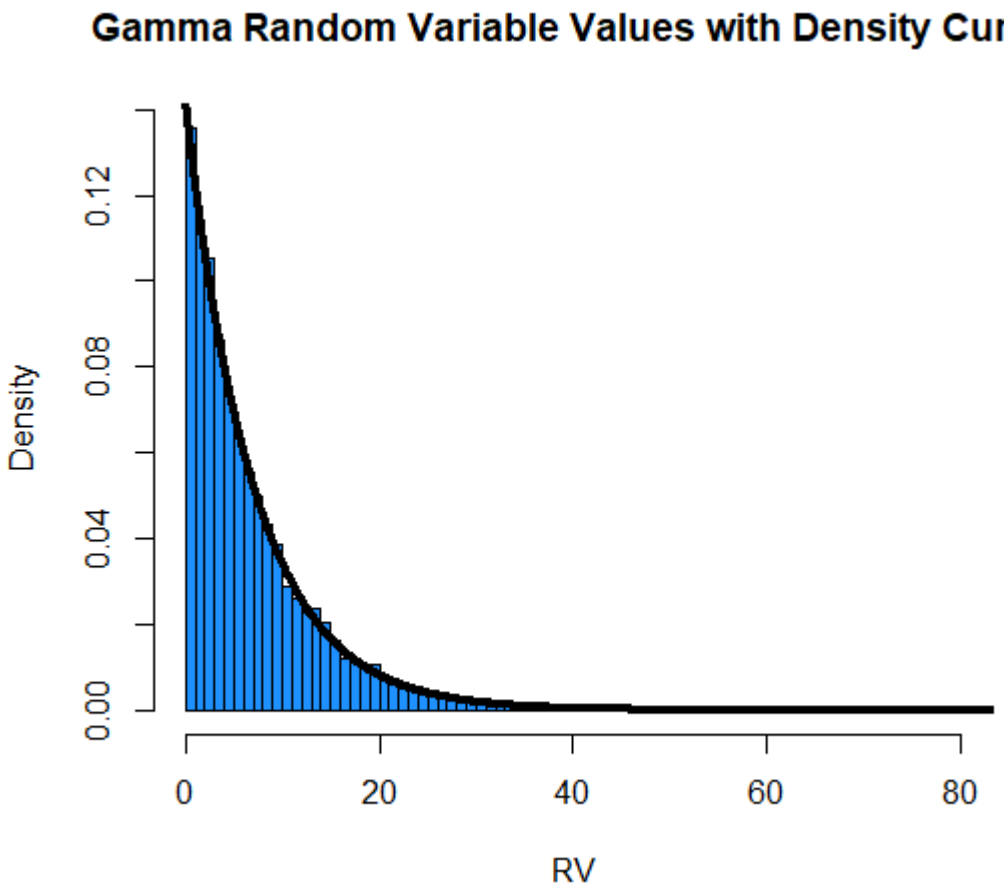


E.

In the second trial I had 11.56 percent of the values below 4. This seems a bit high in comparison to our expected value of 11.26. Looking at the plot it seems I may have a few outliers that are throwing this off. This is about the same amount of error in the first trial with 1000 points. However, This one was off on the high side rather than the low side.

F.

When running the gamma simulation with Alpha set to one I find that the shape of the data looks like it is exponential. The probability density starts off high and gets lower and lower over time.



G.

The above simulation is the Exponential Distribution. As noted before the shape looks like it fits the exponential model. This is because the exponential distribution is the same as the Gamma distribution when alpha is one.

Below is the expression that represents the probability density function.

$$f_x = \begin{cases} \frac{1}{7}e^{-\frac{1}{7}x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$