BU

BOSTON UNIVERSITY

METROPOLITAN COLLEGE

DEPARTMENT OF ADMINISTRATIVE SCIENCES

AD 616: Enterprise Risk Analytics- Aravind hanumantharao

Assignment 3

15-3-2023

What to submit?

Please submit (i) a word file explaining in detail your answers to each question (you can use screenshots of the R to explain your answers) AND (ii) an R file with a separate tab for each question. For each question, make sure you develop the model and present the simulation results – the R file should be self-explanatory. **The assessment of your work will include both the accuracy and the clarity of your word file and the R file. But even if you are struggling with R, we will grade favorably if you demonstrate your understanding of the concepts and how you are planning to solve the problem.**

1. GWS is a company that markets outboard motorboats directly to consumers for recreational use. Recently, they've been developing a project they think has a lot of potential: the first mass market boats with electric motors. They haven't started advertising their new product yet, nor have they organized a presale because they don't want to lose their first-mover advantage. As a result, GWS has a limited understanding of the size of the market for their new project. They plan to retail their boats for $150,000, but after two years, when competition enters the market and the novelty factor wears off, they'll have to drop the price to $70,000. They hire a consultant who estimates that at this price point, over the next two years, demand for the new boats will be somewhere between 2,000 and 15,000, with probabilities as in the table below:

Demand	Probability
2,000-5,000	35%
5,001-10,000	40%
10,001-14,000	20%
14,001-15,000	5%

The fixed cost of manufacturing any number of boats is normally distributed, with a mean of $300 million and a standard deviation of $60 million. They estimate that the variable cost to produce one boat will be a minimum of $77 thousand and a maximum of $100 thousand, with a most likely value of $90,000. Develop four Monte Carlo Simulations to calculate their total profit over the two year period assuming they produce

a. 4,000 boats
b. 8,000 boats
c. 12,000 boats

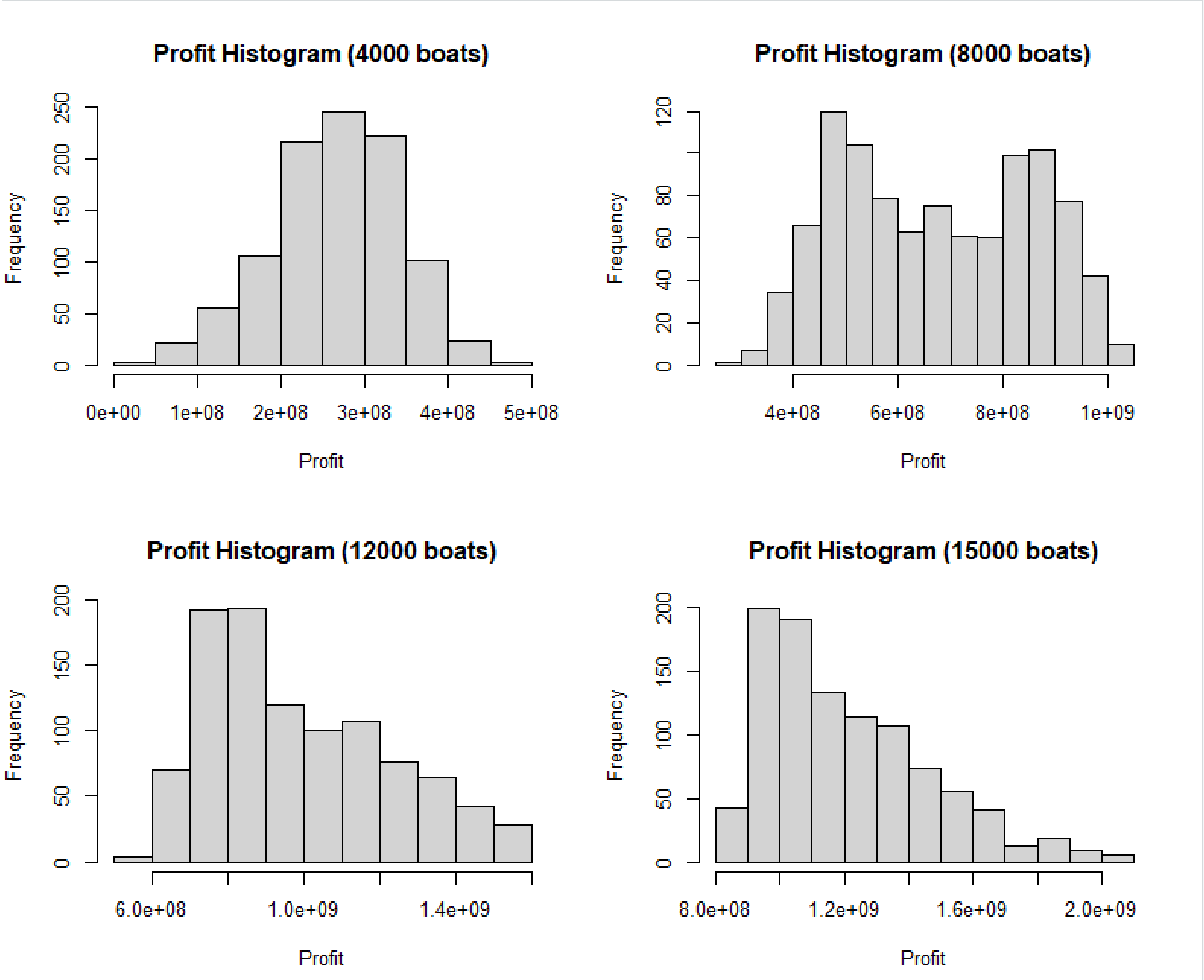
d. 15,000 boats

For each simulation, provide the mean and standard deviation as well as a histogram.

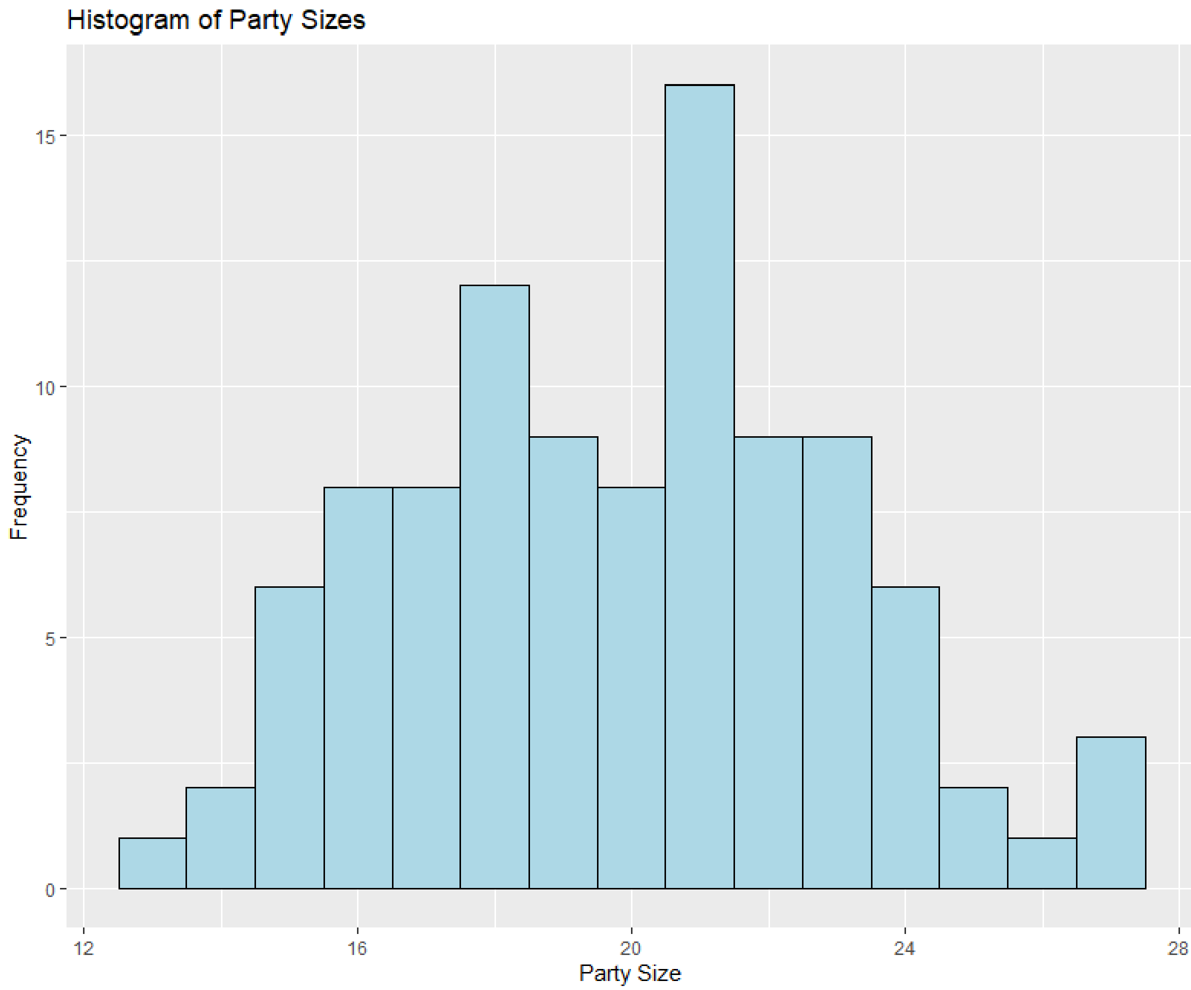
```
#Mean and standard deviation calculation
mean_4000<-mean(profit_4000)
sd_4000<-sd(profit_4000)

mean_8000<-mean(profit_8000)
sd_8000<-sd(profit_8000)

mean_12000<-mean(profit_12000)
sd_12000<-sd(profit_12000)

mean_15000<-mean(profit_15000)
sd_15000<-sd(profit_15000)
```

mean_12000	990926785.491073
mean_15000	1208398511.74443
mean_4000	265339024.073426
mean_8000	673955235.870521

sd_12000	238052438.029153
sd_15000	252291483.346289
sd_4000	76605624.4999069
sd_8000	181494070.900025



2. The Baker's Inn is a small chain of restaurants in the New England area, that, in addition to accepting reservations, allow patrons to reserve an event space for private functions. Until now, the chain has limited reservations for its event spaces to parties of ten or more. The proprietor of the chain is considering expanding the spaces or possibly ceasing to offer private functions, but wishes to model the revenue they can generate. The chain collects data on the previous 100 private functions, including the number of people in each party and the amount each party spent per person (see .csv file).
   a. Create a histogram of the party sizes. Which distribution do you think would best fit the data? Choose from among discrete uniform, geometric, and Poisson.



Based on the histogram of the party sizes, I think that Poisson distribution would be the most appropriate fit for this dataset.

b. In order to best fit the data, we need to shift the distribution to the left by subtracting a positive integer from the value of party size. What number should we subtract to get the best fit? What are the parameter(s) of the best fit distribution?

```r

#b

library(DescTools)
mode_val <- Mode(part_sizes$party_size)
shifted_data <- data.frame(party_size = (part_sizes$party_size - mode_val))

library(MASS)

# Shift the party sizes by subtracting 21
party_sizes_shifted <- part_sizes


party_sizes_shifted$party_size <- party_sizes_shifted$party_size

library(MASS)
library(fitdistrplus)



```

mode_val	int 21

21 is appearing the most and this value needs to be subtracted to the left , to best fit the data

What are the parameters –

```r
#b
# Fit normal distribution
fit_norm_party <- fitdist(party_sizes_shifted$party_size, "norm")
summary(fit_norm_party)


# Subset the data to only include positive values
part_sizes_pos <- subset(part_sizes, party_size > 0)

# Fit gamma distribution
fit_gamma_party <- fitdist(part_sizes_pos$party_size, "gamma")
summary(fit_gamma_party)



# Fit lognormal distribution
fit_lnorm_party <- fitdist(part_sizes_pos$party_size, "lnorm")
summary(fit_lnorm_party)
```

Looking at the summary –

Gamma -had the lowest AIC

```
Fitting of the distribution ' gamma ' by maximum likelihood
Parameters :
       estimate Std. Error
shape 39.261769  5.5289604
rate   1.979026  0.2804762
Loglikelihood:  -256.3037   AIC:  516.6074   BIC:  521.8178
Correlation matrix:
          shape      rate
shape 1.0000000 0.9936395
rate  0.9936395 1.0000000
```

LNORM-

```
Fitting of the distribution ' lnorm ' by maximum likelihood
Parameters :
         estimate Std. Error
meanlog 2.9749092 0.01610418
sdlog   0.1610418 0.01138540
Loglikelihood:  -256.7757   AIC:  517.5513   BIC:  522.7617
Correlation matrix:
        meanlog sdlog
meanlog       1     0
sdlog         0     1

>
```

NORM-

```
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 19.840000  0.3142356
sd    3.142356  0.2221980
Loglikelihood:  -256.3911   AIC:  516.7823   BIC:  521.9926
Correlation matrix:
     mean sd
mean    1  0
sd      0  1
```

We need to choose gamma as our parameter because it has the lowest AIC.

c. We now need to fit a distribution to the per person spending. Which of the following distributions results in the best fit, according to the KS statistic: Cauchy, gamma, logistic, lognormal, normal, or Weibull? What are the parameter(s)?

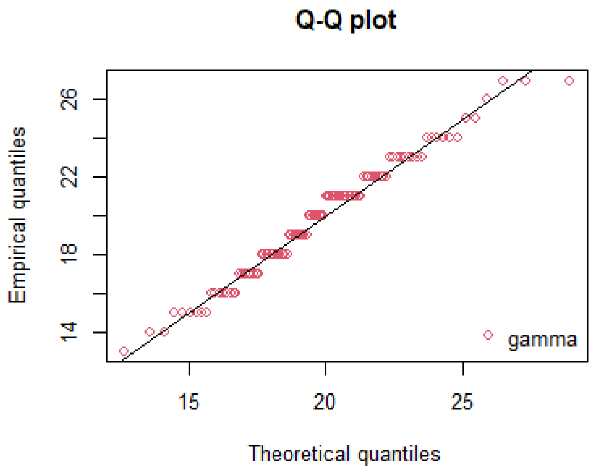
```
#c
# Fit normal distribution
fit_norm_rev <- fitdist(party_sizes_shifted$rev_per_person, "norm")
summary(fit_norm_rev)
gofstat(fit_norm_rev)

# Subset the data to only include positive values
part_sizes_pos <- subset(part_sizes, party_size > 0)

# Fit gamma distribution
fit_gamma_rev <- fitdist(part_sizes_pos$rev_per_person, "gamma")
summary(fit_gamma_rev)
gofstat(fit_gamma_rev)

# Fit lognormal distribution
fit_lnorm_rev <- fitdist(part_sizes_pos$rev_per_person, "lnorm")
summary(fit_lnorm_rev)
gofstat(fit_lnorm_rev)

# Fit cauchy distribution
fit_cauchy_rev <- fitdist(party_sizes_shifted$rev_per_person, "cauchy")
summary(fit_cauchy_rev)
gofstat(fit_cauchy_rev)


fit_logistics_rev <- fitdist(party_sizes_shifted$rev_per_person, "logis")
summary(fit_logistics_rev)
gofstat(fit_logistics_rev)

fit_weibull_rev <- fitdist(part_sizes_pos$rev_per_person, "weibull")
summary(fit_weibull_rev)
gofstat(fit_weibull_rev)
```

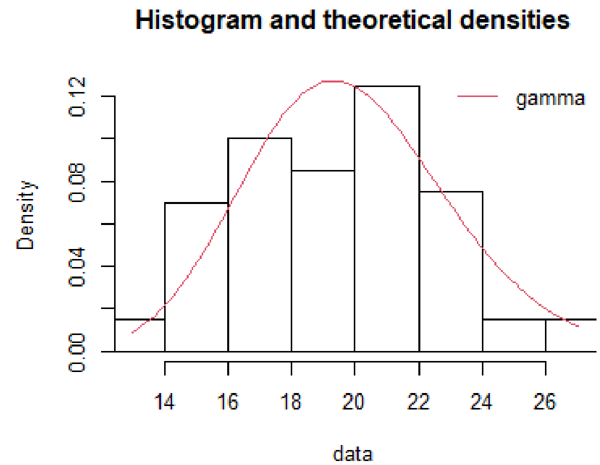
Conducting all the k- test , I came to a conclusion –

```
> gofstat(fit_gamma_rev)
Goodness-of-fit statistics
                             1-mle-gamma
Kolmogorov-Smirnov statistic  0.03600846
Cramer-von Mises statistic    0.01748614
Anderson-Darling statistic    0.15435859

Goodness-of-fit criteria
                               1-mle-gamma
Akaike's Information Criterion    1037.958
Bayesian Information Criterion    1043.169
>
```

The k test proves that , gamma had the lowest value and I need to choose that parameter.

d. Create a QQ plot and a density comparison plot for the distribution you chose from part (c). Does the distribution appear to be a good fit?





Yes the plot is a good fit and seen the points are closer to the line and the distribution is normally distributed towards the centre.

e. Does it appear that party size and per person spending are correlated?

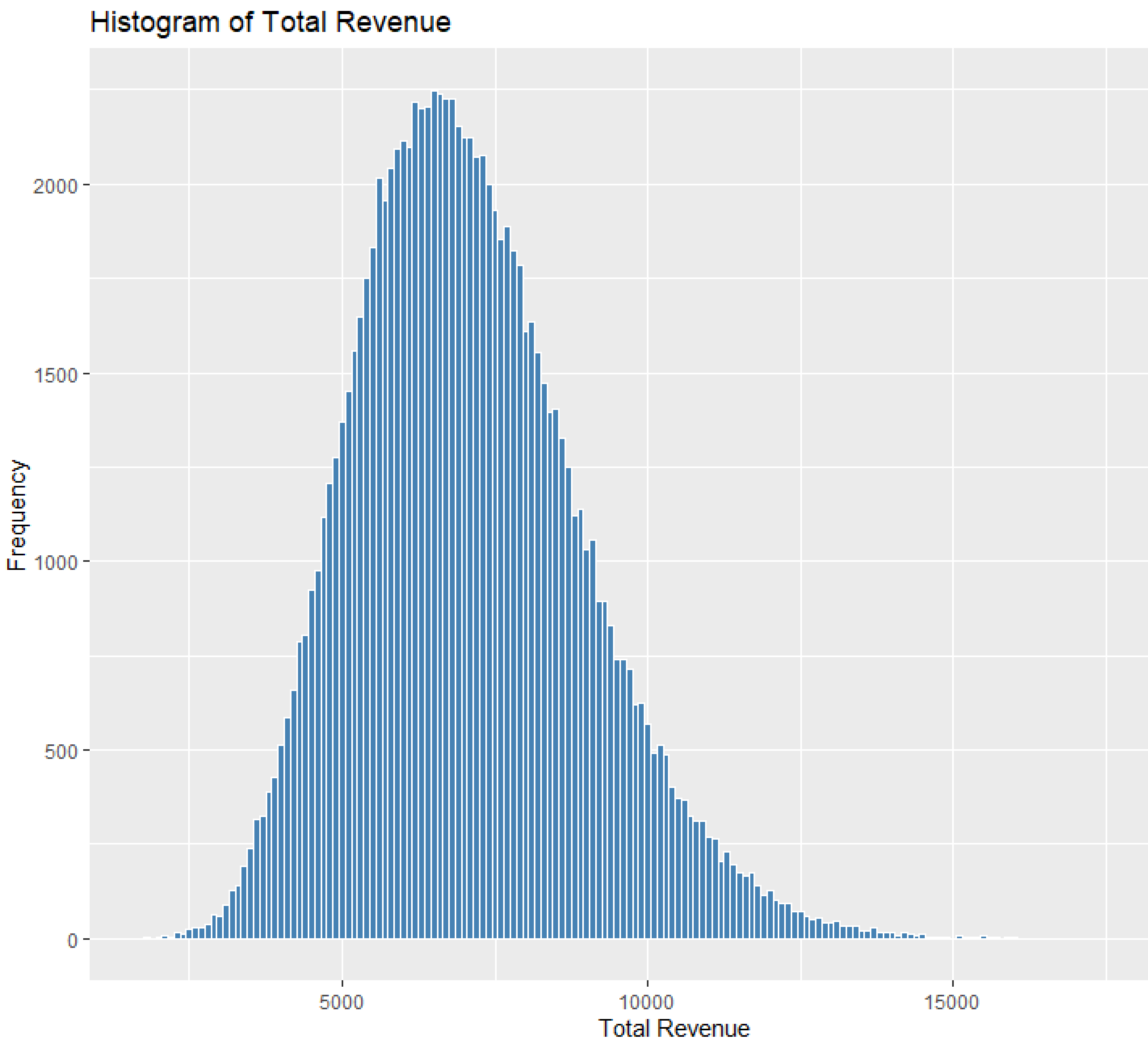
```
> cor(party_sizes_shifted$party_size,part_sizes$rev_per_person)
[1] 0.5068473
>
```

Yes it is moderately positively correlated .

f. Using the distributions and parameters from parts (a)-(c) and the correlation from part e, construct an MC simulation with 100,000 trials to create a risk profile for the revenue generated from a single event. Create a histogram and provide the mean and standard deviation. (Hint: don't forget to shift the distribution of party size back to the right.)

```r
#Monte-Carlo
# Set seed for reproducibility
set.seed(123)


mean_party_sizes <-mean(part_sizes$party_size)
mean_rev_per_person <-mean(part_sizes$rev_per_person)

# Number of trials
n <- 100000

# Correlation matrix
corr_mat <- matrix(c(1, corr, corr, 1), ncol = 2)

# Simulate correlated gamma distributions for party size and revenue per person
sim_data <- MASS::mvrnorm(n, mu = c(mean_party_sizes, mean_rev_per_person), Sigma = corr_mat)
partysize <- rgamma(n, shape = fit_gamma_party$estimate[1], rate = fit_gamma_party$estimate[2]) + 21
rev_person <- rgamma(n, shape = fit_gamma_rev$estimate[1], rate = fit_gamma_rev$estimate[2])

# Calculate total revenue for each trial
total_revenue <- partysize * rev_person

# Calculate mean and standard deviation of total revenue
mean_total_revenue <- mean(total_revenue)
sd_total_revenue <- sd(total_revenue)

# Create histogram of total revenue
ggplot(data.frame(total_revenue), aes(x = total_revenue)) +
  geom_histogram(binwidth = 100, fill = "steelblue", color = "white") +
  labs(x = "Total Revenue", y = "Frequency", title = "Histogram of Total Revenue")
```

mean_total_revenue	7077.42117589646
sd_total_revenue	1854.02792739739



g. According to your simulation, how often will a private function generate at least $5000?

```
#g
# Proportion of trials where total revenue is at least $5000
prop_revenue_5000 <- mean(total_revenue >= 5000)

# Print proportion
prop_revenue_5000

```

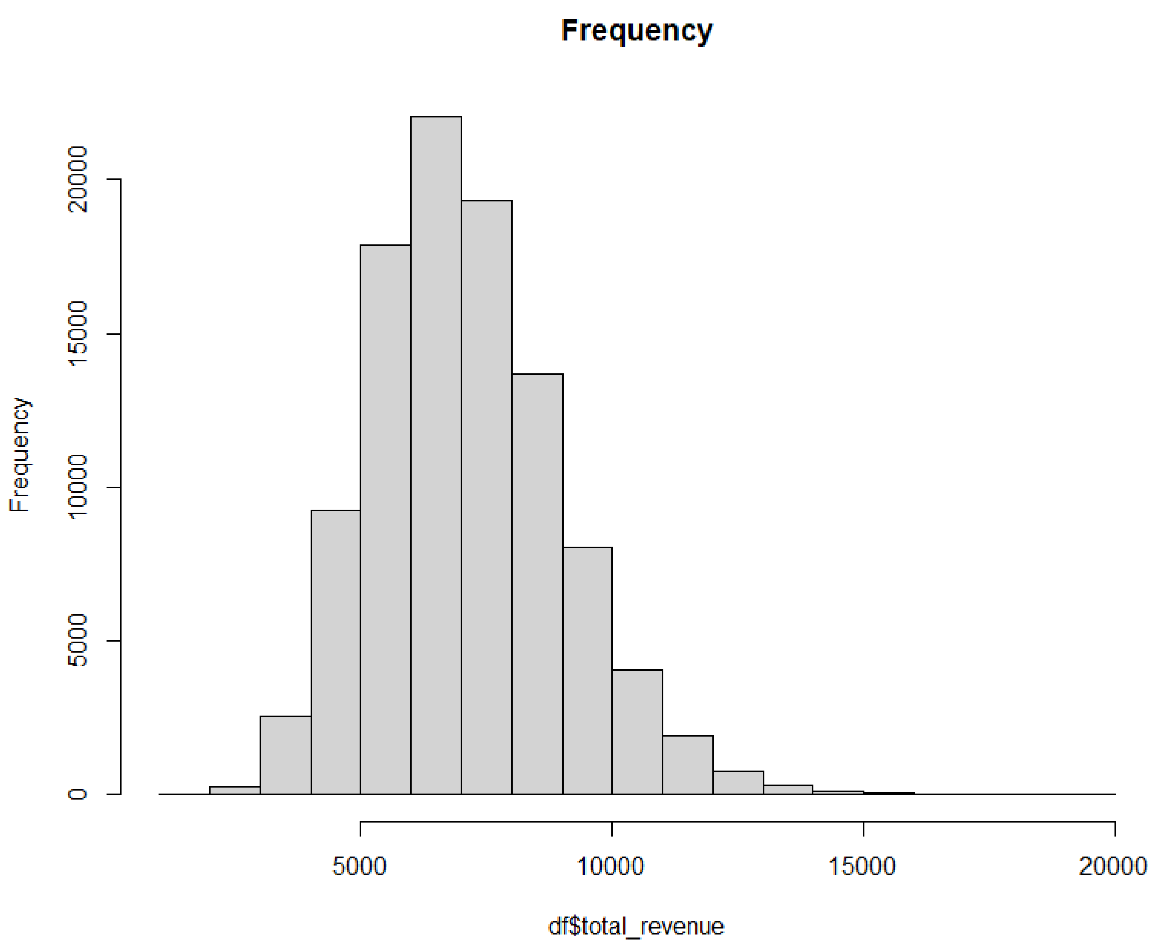
```
> prop_revenue_5000
[1] 0.87991
>
```

87 percent of the times the values are above 5000.

h. Repeat parts (f) and (g), but build a simulation that ignores the correlation. How does this affect your results?

```
#h
#Creating a monte carlo simulation for 100,000 trails using the previous parameters

n<-100000
df<-data.frame(party_size=rgamma(n,fit_gamma_party$estimate[1],fit_gamma_party$estimate[2]) %>% round()+21,
               rev_per_person=rgamma(n,fit_gamma_rev$estimate[1],fit_gamma_rev$estimate[2]))



df$total_revenue<-df$party_size*df$rev_per_person

hist(df$total_revenue,main="Frequency")

mean(df$total_revenue)
sd(df$total_revenue)
```

```
> mean(df$total_revenue)
[1] 7090.207
> sd(df$total_revenue)
[1] 1865.931
>
```



Not having the correlation , the mean and the standard deviation increases from before . That's the only change from the correlation equation.

AD 616: Enterprise Risk Analytics

Assignment 1

Aravind Hanumantha Rao

What to submit?

Please submit (i) a word file explaining in detail your answers to each question (you can use screenshots of the R to explain your answers) AND (ii) an R file with a separation for each question. For each question, make sure you develop the model and present the simulation results – the R file should be self-explanatory. **The assessment of your work will include both the accuracy and the clarity of your word file and the R Code.**

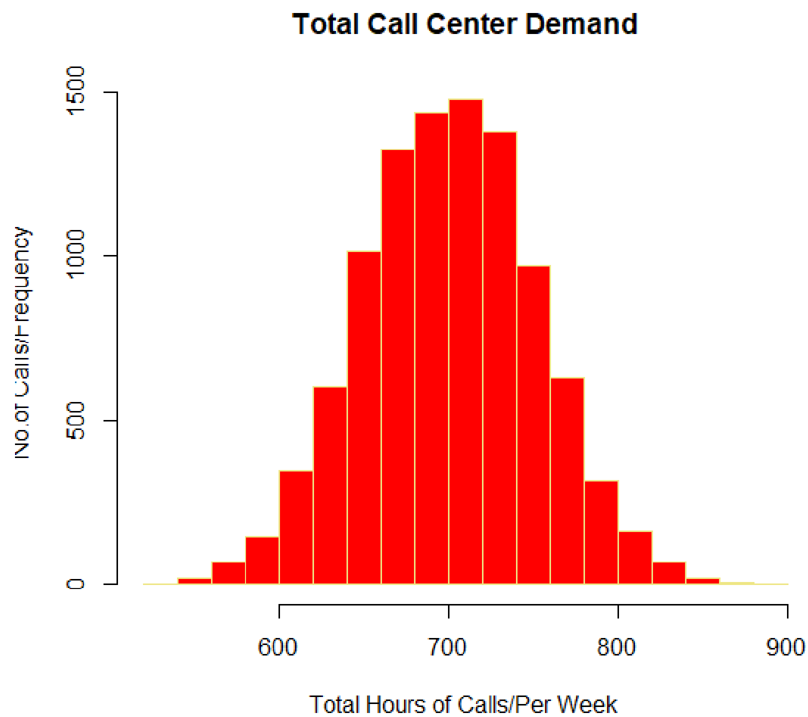
1. Consider a call center that receives its demand over a set of different travel websites. The weekly demand for each website is normally distributed with a mean and standard deviation given in Table 1. Develop a R script that creates a simulation with 100,000 trials to determine total call center demand.

Table 1: Weekly demand of travel sites (in hours)

Travel **Site**	Mean	Standard Deviation
A	200	20
B	50	10
C	100	15
D	150	30
E	100	30
F	100	10

a) What are the mean and standard deviation of total call center demand according to your simulation?
Mean and standard deviation is determined from the total call center from all the random normal variables. total call center is calculated after normally distributing all the variables and adding up all the travel sites.

b) Develop a histogram that models the risk profile for total call center demand.


Total Call Center Demand
No.of Calls/Frequency
0
500
1000
1500
600
700
800
900
Total Hours of Calls/Per Week

2. A cell phone manufacturer is considering to offer a refund to its customers whose battery fails before 5 years. The refunds start at $10, and increase by $1.50 for every month the battery falls short of 5 years. For example, a customer whose battery fails after 4 years and 6 months would receive a refund of $19. A customer whose battery fails after 5 years would receive no refund. Previous studies show that a battery's life is normally distributed with a mean of 7 years and standard deviation of 2 years. Develop a simulation with 100,000 trials for the amount of a refund.

   a) According to your simulation, what is the expected cost per cell phone to the manufacturer of this offer?

      Replacement is 45.46 . To derive to this calculation , we had to calculate the replacement cost . includes the refund for the phones less than 5 years and the total battery life that falls short of 5 years and divide it by the total trials.(100000)

   b) According to your simulation, what is the probability a refund will be paid?

      15.84 .phones less than 5 years divided by total trials (100000) *100 , gives the probability

   c) According to your simulation, what is the average cost per refund?

      To calculate the average cost per fund , we need to know the replacement cost and phones less than 5 years.

3. A coffee cart opens at 7:00 am, and they generally try to prepare a batch large enough to accommodate their customers until 10:00 am, when the cart closes. The operator only sells 16 oz. servings, and each serving costs $0.50 to prepare, which includes all costs of production and any dairy/sweetener customers may add. Any coffee that isn't sold before 10:00 am is considered stale and disposed of for no monetary gain. When purchased, the coffee is poured into a ripple cup, which costs an additional $0.15 per cup. Demand over this period is normally distributed with a mean of 125 and a standard deviation of 35. Each cup retails for $2.75. The operator of the cart must also purchase a municipal license, which costs a flat $100 for the three hours. Assume there are no other costs associated with the cart. Develop a R model with 100,000 trials that simulates the daily profit resulting from the preparation of 75, 100, 120, 140, 160, and 180 servings of coffee a day (run them one at a time).

   a) For each option, what is the expected profit, and which option results in the highest expected profit?

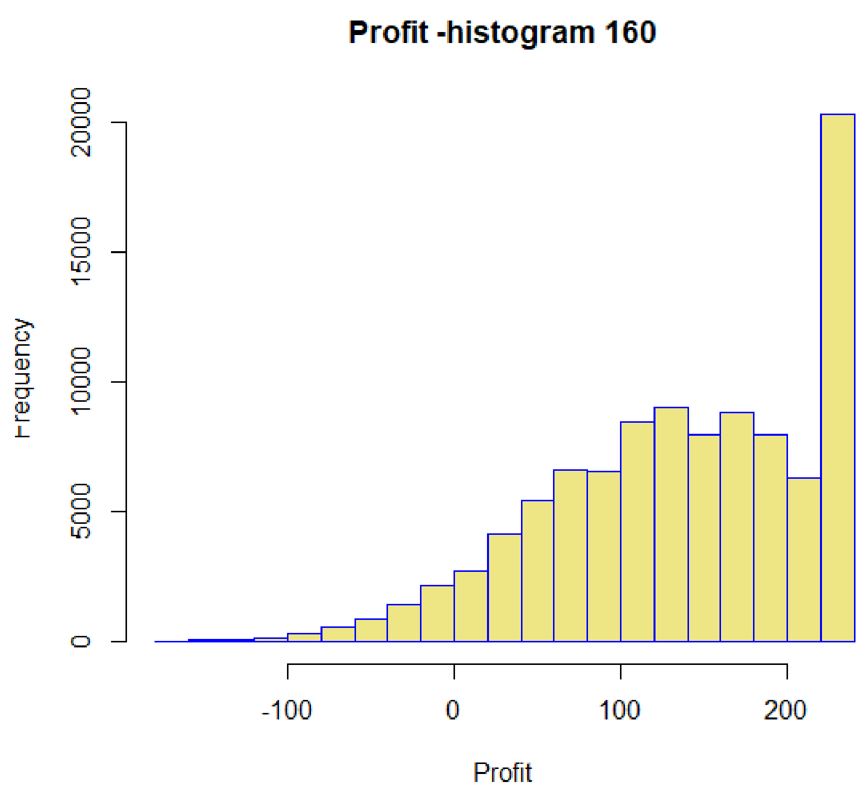
	Cups_Per_Day	Profit_Per_Day
1	75	54.34610
2	100	97.10033
3	120	121.43658
4	140	134.30889
5	160	136.58926
6	180	131.77223

To derive to the expected profit the following measures had to be taken place to get the values.

1) Cost of the cup is 0.15 +0.5 and the profit of the cup or margin is 2.75-0.65
2) Rent is 100
3) Now for each serving of coffee we need to calculate whether they make a profit or not.
4) Pmin because the minimum they can sell in that day depends upon the servings of the coffee run on the day.

The conditions that had to be taken place is whether the customers were served with the ripple cup or served without it . the 2 conditions that I have kept is the total money they make from minusing the cost of cup and servings with the rent. The other one is without the cup and therefore only the cost of production is taken into consideration. This would give the expected profits for all the coffee servings and help in identifying the profit_per_day

b) Create a histogram that displays the risk profile of profit for the number of servings with the highest expected profit.



As seen from the table , the highest profit was from 160 servings per day.

BOSTON UNIVERSITY

METROPOLITAN COLLEGE

DEPARTMENT OF ADMINISTRATIVE SCIENCES

AD 616: Enterprise Risk Analytics

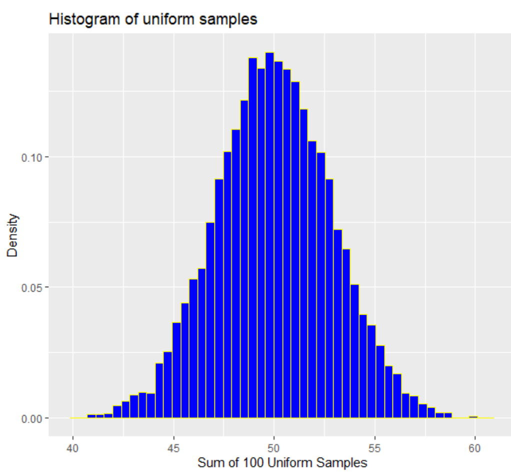
Assignment 2

Aravind Hanumantha rao

1) Answer the following two questions.

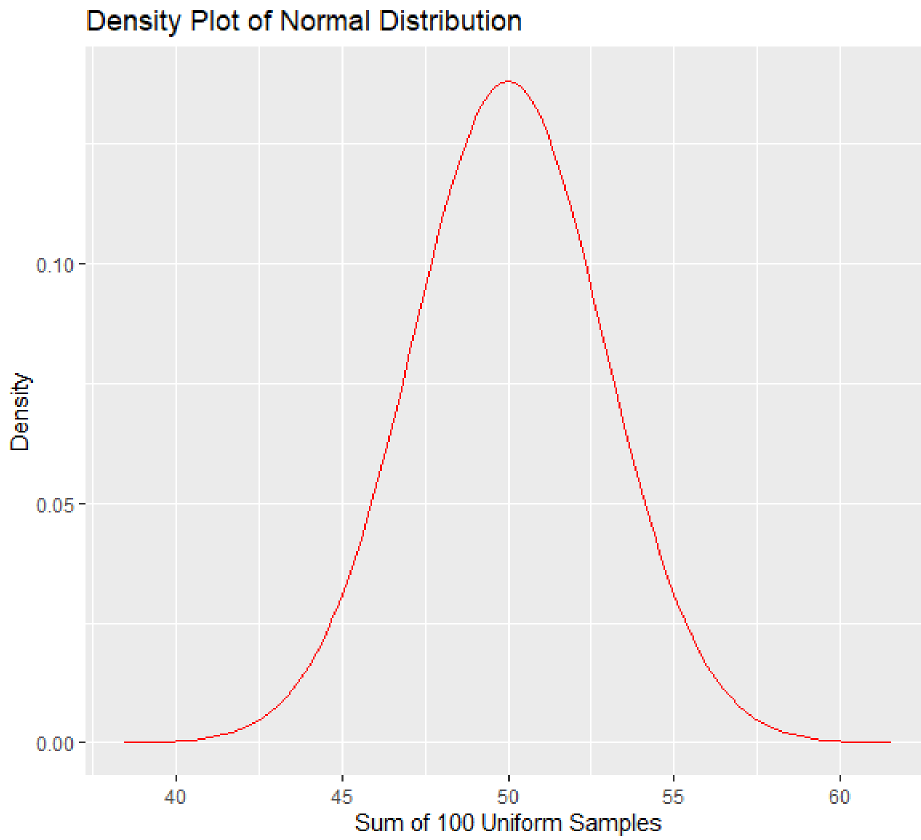
   a) According to the *central limit theorem*, the sum of *n independent identically distributed* random variables will start to resemble a normal distribution as *n* grows large. The mean of the resulting distribution will be *n* times the mean of the summands, and the variance *n* times the variance of the summands. Demonstrate this property using Monte Carlo simulation. Over 10,000 trials, take the sum of 100 uniform random variables (with min=0 and max=1). Note: the variance of the uniform distribution with min 0 and max 1 is 1/12. Include:

   i) A histogram of the results of the MC simulation



The histogram shows the sum of 100 uniform samples when run for 10,000 trials . As the histogram is formed for the data .
So the sum of 100 samples is formed from the runif function and entered into the matrix .
Summing the rows of 10,000 trails with 100 columns gives the sum of the samples

   ii) A density plot of a normal distribution with the appropriate mean and standard deviation



As we can see from the histogram and relate with the density plot , the mean is calculated from the average of min and max of the uniform distribution . If the min is 0 and max is 1 then the average of the uniform distribution is 0.5 because it's a uniform distribution . When multiplied with the 100 uniform variables , it gives a mean of 50. Knowing the variance of the uniform distribution is 1/12 , the 100 uniform variables will have a variance of 8.33.

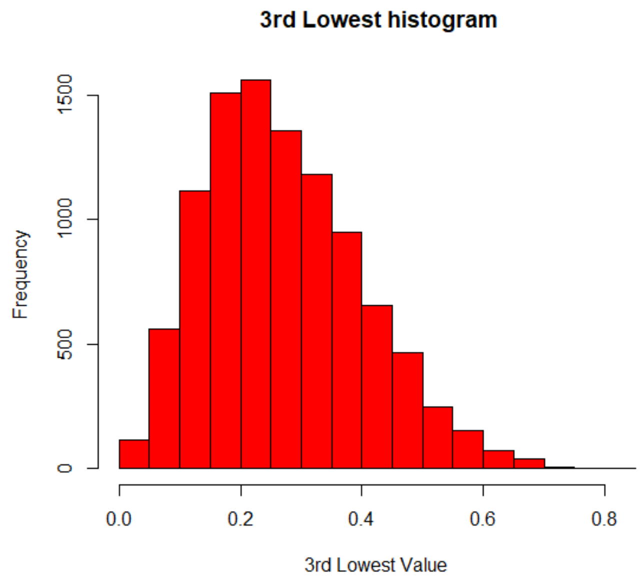
So the mean being 50 and variance of 8.3 and using 4 standard deviations which covers a area of 99.75 percent from the mean, the density plot is constructed.

iii) The mean and standard deviation of the MC simulation.

```
> # The mean and standard deviation of the MC simulation.
> mean(s_samples)
[1] 49.98811
> sd(s_samples)
[1] 2.884825
```

b) According to probability theory, if $\{X_1, \ldots X_n\}$ are independent and *uniform distributed* random variables with min=0 and max=1, then the k^{th} lowest will follow a *beta distribution* with parameters *shape1=k*, *shape2=n+1-k.* Demonstrate this property using Monte Carlo simulation. Simulate 10,000 trials. For each trial, generate 10 uniform random variables and select the 3rd lowest. Include:

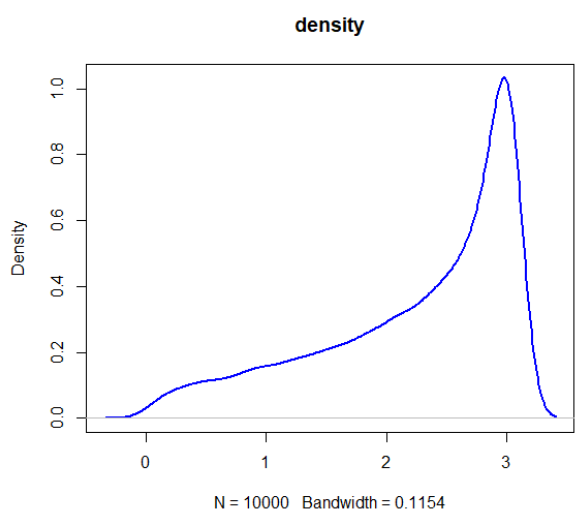
i) A histogram of the results of the MC simulation



When applying the sort function for samples with the given attributes and running the runif function with the trials and the n value we get this histogram .

Sort function with margin one, sorts by the rows in ascending order and assigning k = 3 will take the 3rd lowest from the sorted values from the 3rd column. They consist of 10 columns and 10,000 rows.

ii) A density plot of the beta distribution with the appropriate parameters



Given the 2 parameters and the k_lowest values we can construct the beta density. calculated the values of k_lowest from the sorting function . These values get inserted into the S1 and S2 parameters to give a beta distribution

iii) The mean and standard deviation of the MC simulation

```
> mean(k_lowest)
[1] 0.2715974
> sd(k_lowest)
[1] 0.1281659
```

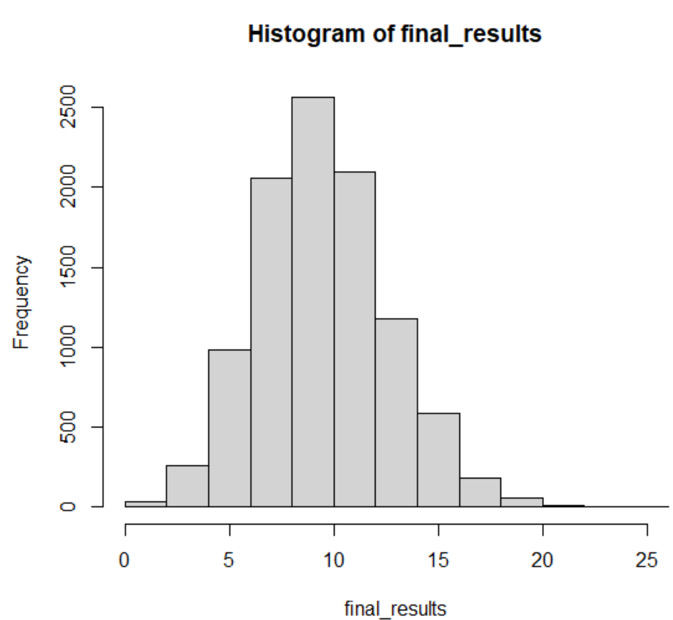
2) Assume you represent a worldwide distributor of classic cars. Create a Monte Carlo simulation with 10,000 trials to demonstrate the property that, if the amount of time it takes before your next customer makes a purchase can be modelled using an exponential distribution with a rate of 10 per day, then the number of times customers will make purchases in a day will follow a Poisson distribution with $\lambda = 10$. Develop a histogram to reinforce your result.-

Given the rate of 10 and trials of 10000, we can calculate the random exponential values . These values can be appended into the matrix. The reason I took 100 columns is because to get a dispersed values for the time taken for the next customers. The columns are for each time and the rows are the purchases for each trial.

So , the apply function sums the exponentials to get the know the value of one purchase and that is transposed into a row and column matrix . the reason of transpose is because it only cumsums by the rows and not to the columns and transpose it can fix the issue. Transpose converts the rows of the matrix in column and column of the matrix in rows.

Then for the FOR loop will check until the values reach 1 and count it as one purchase for the day and it does it for the other 10,000 trials. It sums the results to produce a final result of the purchase for the rows and the columns.

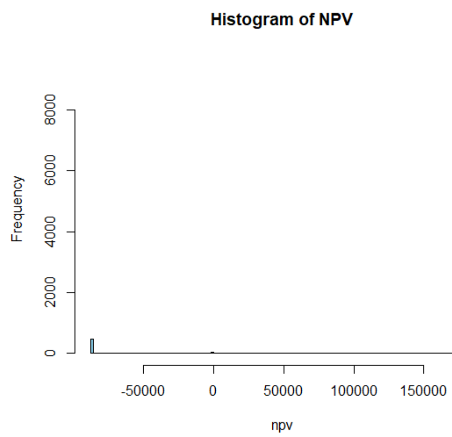
This can be produced into a histogram:-



3) A life insurance company is pricing a new policy to sell to a group of 45-year-old male non-smokers. They determine that the probability that a member of this group will die *X* years from the day they purchase the policy can be modeled with a Weibull distribution with shape parameter 4.5 and scale parameter 39, measured in years. The term of the policy is 20 years. At the end of every month, policy holders are expected to pay $115. If a policy holder in good standing dies during the term of the policy, his beneficiaries receive a lump sum of $1,000,000 at the end of the month. Every month there is a 0.3% chance that the policy holder will let the policy lapse (i.e. he will permanently stop paying premiums and forfeit his right to the benefit). The insurance company calculates cost of funds using a rate of 6.5%

Create a Monte Carlo simulation with 10,000 trials of the above scenario to calculate the net present value of cash flows to the insurance company for one policyholder.

a) Create a histogram describing the NPV. How would you characterize the distribution?



It shows 3 conditions – one is 185093 , zero and -86923

b) What are the mean and standard deviation of the NPV? On balance, is the insurance company making a profit?
Mean is 171470.63 and standard deviation is 58895.46

c) Provide a 95% confidence interval for the mean of the NPV. Interpret the result.
95% confidence interval for the mean of the NPV (170316.167266759, 172625.106389196).

d) How many iterations would be necessary to provide a **99%** confidence interval with a half width of $200?
The number of iterations needed for a 99% confidence interval with a half width of $200 is 575358

e) The company can be 90% sure their npv will be at least *x.* Solve for *x.*
The company can be 90% sure their NPV will be at least $185092.51.
The company can be 99% sure their npv will be at least *y.* Solve for *y.*
The company can be 99% sure their NPV will be at least $-86923.63.

Now assume the insurance company underwrites 1,000 policyholders. Create a Monte Carlo simulation with 1,000 trials to calculate the net present value of cashflows for the insurance company made to all the policy holders. *(Hint: Recycle your work above. Create a list of data frames, where each element of the list represents one trial.)* Answer questions (a)-(e) above under this assumption.