MET AD 616 Assignment 3

JingjianGao

2023-11-21

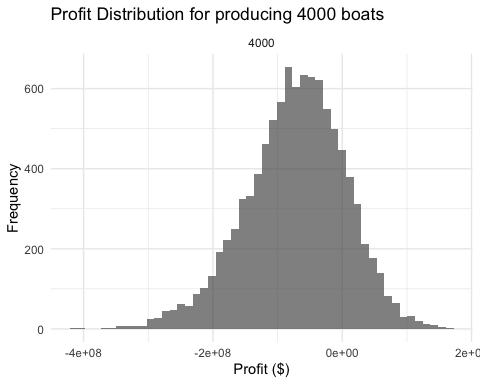
## Question #1

### Monte Carlo Simulation

library(ggplot2)  
  
set.seed(100)  
mean\_fixed\_cost <- 300e6   
std\_dev\_fixed\_cost <- 60e6   
retail\_price\_initial <- 150000 # Initial retail price  
retail\_price\_final <- 70000 # Final retail price after two years  
min\_variable\_cost <- 77000   
most\_likely\_variable\_cost <- 90000   
max\_variable\_cost <- 100000   
n\_simulations <- 10000  
  
# The Demand ranges and their probabilities  
demand\_ranges <- list(c(2000, 5000), c(5001, 10000), c(10001, 14000), c(14001, 15000))  
probabilities <- c(0.35, 0.40, 0.20, 0.05)  
  
  
simulation\_demand <- function() {  
 index <- sample(1:length(demand\_ranges), size = 1, prob = probabilities)  
 return(sample(demand\_ranges[[index]][1]:demand\_ranges[[index]][2], size = 1))  
}  
  
simulation\_profit <- function(n\_boats) {  
 fixed\_costs <- rnorm(n\_simulations, mean\_fixed\_cost, std\_dev\_fixed\_cost)  
 variable\_costs <- runif(n\_simulations, min\_variable\_cost, max\_variable\_cost)  
 profits <- numeric(n\_simulations)  
   
 for (i in 1:n\_simulations) {  
 demand <- simulation\_demand()  
 units\_sold\_at\_high\_price <- min(n\_boats, demand)  
 units\_sold\_at\_low\_price <- max(n\_boats - demand, 0)  
 total\_variable\_cost <- variable\_costs[i] \* n\_boats  
 total\_cost <- fixed\_costs[i] + total\_variable\_cost  
 total\_revenue\_high <- retail\_price\_initial \* units\_sold\_at\_high\_price  
 total\_revenue\_low <- retail\_price\_final \* units\_sold\_at\_low\_price  
 profits[i] <- total\_revenue\_high + total\_revenue\_low - total\_cost  
 }  
   
 return(profits)  
}

# 4000 boats

n\_boats <- 4000  
results <- data.frame()  
  
profits <- simulation\_profit(n\_boats)  
mean\_profit <- mean(profits)  
std\_dev\_profit <- sd(profits)  
results <- rbind(results, data.frame(Production = n\_boats, Profit = profits, MeanProfit = mean\_profit, StdDevProfit = std\_dev\_profit))  
  
ggplot(results, aes(x=Profit)) +  
 geom\_histogram(bins=50, alpha=0.7) +  
 facet\_wrap(~ Production, scales="free") +  
 labs(title="Profit Distribution for producing 4000 boats",  
 x="Profit ($)",  
 y="Frequency") +  
 theme\_minimal()



mean\_profit

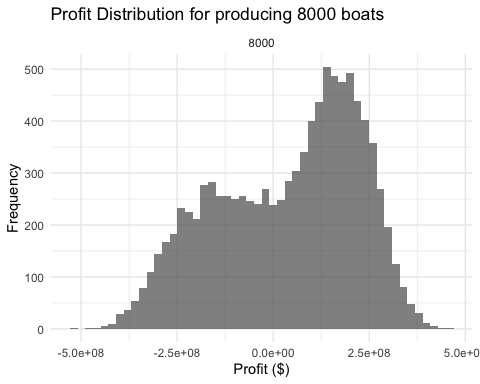
## [1] -72559546

std\_dev\_profit

## [1] 77228476

## 8000 boats

n\_boats <- 8000  
results <- data.frame()  
  
profits <- simulation\_profit(n\_boats)  
mean\_profit <- mean(profits)  
std\_dev\_profit <- sd(profits)  
results <- rbind(results, data.frame(Production = n\_boats, Profit = profits, MeanProfit = mean\_profit, StdDevProfit = std\_dev\_profit))  
  
ggplot(results, aes(x=Profit)) +  
 geom\_histogram(bins=50, alpha=0.7) +  
 facet\_wrap(~ Production, scales="free") +  
 labs(title="Profit Distribution for producing 8000 boats",  
 x="Profit ($)",  
 y="Frequency") +  
 theme\_minimal()



mean\_profit

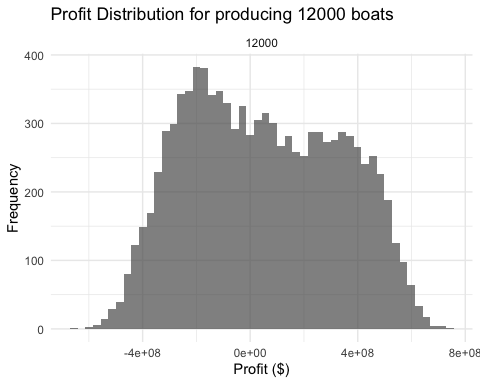
## [1] 37542930

std\_dev\_profit

## [1] 184874898

## 12000 boats

n\_boats <- 12000  
results <- data.frame()  
  
profits <- simulation\_profit(n\_boats)  
mean\_profit <- mean(profits)  
std\_dev\_profit <- sd(profits)  
results <- rbind(results, data.frame(Production = n\_boats, Profit = profits, MeanProfit = mean\_profit, StdDevProfit = std\_dev\_profit))  
  
ggplot(results, aes(x=Profit)) +  
 geom\_histogram(bins=50, alpha=0.7) +  
 facet\_wrap(~ Production, scales="free") +  
 labs(title="Profit Distribution for producing 12000 boats",  
 x="Profit ($)",  
 y="Frequency") +  
 theme\_minimal()



mean\_profit

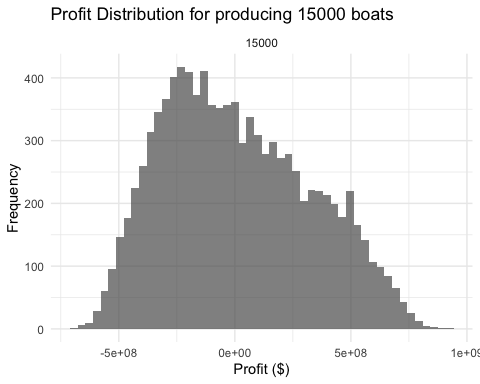
## [1] 49059740

std\_dev\_profit

## [1] 281622662

## 15000 boats

n\_boats <- 15000  
results <- data.frame()  
  
profits <- simulation\_profit(n\_boats)  
mean\_profit <- mean(profits)  
std\_dev\_profit <- sd(profits)  
results <- rbind(results, data.frame(Production = n\_boats, Profit = profits, MeanProfit = mean\_profit, StdDevProfit = std\_dev\_profit))  
  
ggplot(results, aes(x=Profit)) +  
 geom\_histogram(bins=50, alpha=0.7) +  
 facet\_wrap(~ Production, scales="free") +  
 labs(title="Profit Distribution for producing 15000 boats",  
 x="Profit ($)",  
 y="Frequency") +  
 theme\_minimal()



mean\_profit

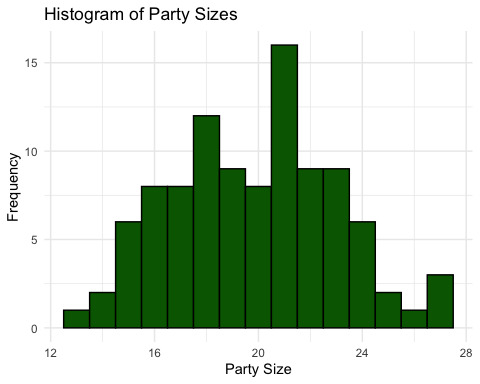
## [1] 9763938

std\_dev\_profit

## [1] 316480485

## Question #2

data <- read.csv("/Users/billg/Desktop/MA 679/MET 616/Module-3/Assignment 3 Problem 2 data.csv")  
  
# a)  
ggplot(data, aes(x=party\_size)) +  
 geom\_histogram(binwidth = 1, fill="darkgreen", color="black") +  
 labs(title="Histogram of Party Sizes", x="Party Size", y="Frequency") +  
 theme\_minimal()



# Since the histogram does not support the idea that each party size is equally likely, the discrete uniform distribution is not a good fit.  
  
# The histogram does not show a clear Poisson distribution as it lacks a distinct peak at lower values.  
  
# The geometric distribution seems to be a good fit since the histogram shows a decline in frequency as party size increase.  
  
# b)  
library(MASS)  
library(survival)  
library(fitdistrplus)  
  
shift\_value <- 10  
shifted\_data <- data$party\_size - shift\_value  
shifted\_data <- shifted\_data[shifted\_data > 0]  
  
fit\_geom <- fitdist(shifted\_data, "geom", discrete = TRUE)  
  
gof\_geom <- gofstat(fit\_geom)  
  
print(fit\_geom)

## Fitting of the distribution ' geom ' by maximum likelihood   
## Parameters:  
## estimate Std. Error  
## prob 0.09225092 0.008788348

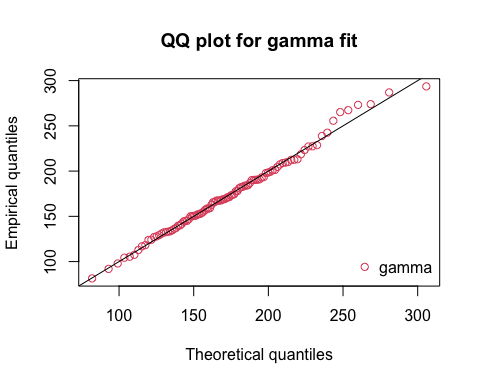
print(gof\_geom)

## Chi-squared statistic: 133.3622   
## Degree of freedom of the Chi-squared distribution: 6   
## Chi-squared p-value: 2.516317e-26   
## the p-value may be wrong with some theoretical counts < 5   
## Chi-squared table:  
## obscounts theocounts  
## <= 5 9.000000 44.050669  
## <= 7 16.000000 9.846613  
## <= 8 12.000000 4.253018  
## <= 9 9.000000 3.860673  
## <= 11 24.000000 6.685750  
## <= 12 9.000000 2.887756  
## <= 13 9.000000 2.621358  
## > 13 12.000000 25.794162  
##   
## Goodness-of-fit criteria  
## 1-mle-geom  
## Akaike's Information Criterion 669.1260  
## Bayesian Information Criterion 671.7311

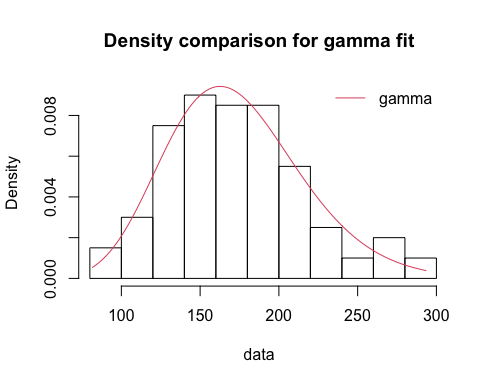
# In this case, we should subtract 10 to get the best fit  
  
# c)  
fit\_gamma <- fitdist(data$rev\_per\_person,"gamma")  
print(summary(fit\_gamma))

## Fitting of the distribution ' gamma ' by maximum likelihood   
## Parameters :   
## estimate Std. Error  
## shape 15.9386656 2.22539298  
## rate 0.0918413 0.01302526  
## Loglikelihood: -516.9792 AIC: 1037.958 BIC: 1043.169   
## Correlation matrix:  
## shape rate  
## shape 1.0000000 0.9842839  
## rate 0.9842839 1.0000000

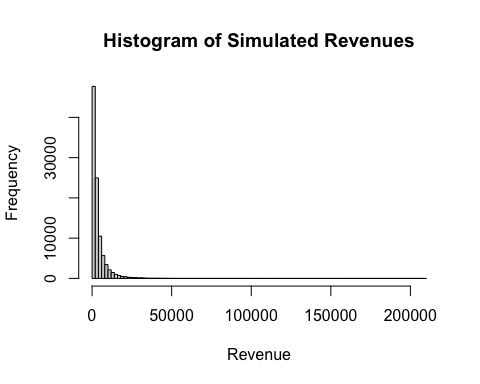
# gamma distribution results in the best fit.  
# The shape parameter is about 15.9386656  
  
# d)  
qqplot <- qqcomp(fit\_gamma, main="QQ plot for gamma fit")



densityplot <- denscomp(fit\_gamma, main= "Density comparison for gamma fit")



# The Gamma distribution appears to be a good fit  
  
# e)  
# It appears that there is a significant positive correlation between party size and per person spending in the data.  
  
# f)  
shift\_value <- 10   
shape\_param <- 12.5432   
scale\_param <- 12.3414   
correlation\_coefficient <- 0.507   
n <- 100000  
  
party\_size\_simulation <- rgeom(n, prob = 1 / (mean(data$party\_size) - shift\_value + 1)) + shift\_value  
  
spending\_simulation <- rgamma(n, shape = shape\_param, scale = scale\_param)  
  
spending\_simulation\_adjusted <- spending\_simulation \* (1 + correlation\_coefficient \* (party\_size\_simulation - mean(party\_size\_simulation)) / sd(party\_size\_simulation))  
  
revenues <- party\_size\_simulation \* spending\_simulation\_adjusted  
  
hist(revenues, breaks = 100, main = "Histogram of Simulated Revenues", xlab = "Revenue")



mean\_revenue <- mean(revenues)  
std\_dev\_revenue <- sd(revenues)  
  
mean\_revenue

## [1] 3881.866

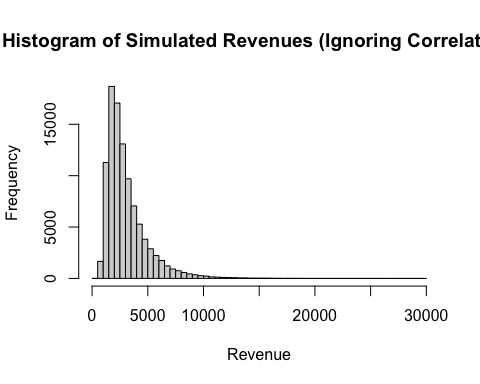
std\_dev\_revenue

## [1] 5458.55

# g)  
over5000 <- sum(revenues >= 5000)  
frequency\_over\_5000 <- over5000/n  
frequency\_over\_5000

## [1] 0.21136

# About 21% of the time  
  
# f)  
shift\_value <- 10   
shape\_param <- 12.5432   
scale\_param <- 12.3414   
n <- 100000  
  
  
party\_size\_simulation\_independent <- rgeom(n, prob = 1 / (mean(data$party\_size) - shift\_value + 1)) + shift\_value  
  
spending\_simulation\_independent <- rgamma(n, shape = shape\_param, scale = scale\_param)  
  
  
revenues\_independent <- party\_size\_simulation\_independent \* spending\_simulation\_independent  
  
hist(revenues\_independent, breaks = 100, main = "Histogram of Simulated Revenues (Ignoring Correlation)", xlab = "Revenue")



mean\_revenue\_independent <- mean(revenues\_independent)  
std\_dev\_revenue\_independent <- sd(revenues\_independent)  
  
over\_5000\_independent <- sum(revenues\_independent >= 5000)  
frequency\_over\_5000\_independent <- over\_5000\_independent / n  
  
mean\_revenue\_independent

## [1] 3069.242

std\_dev\_revenue\_independent

## [1] 1869.811

frequency\_over\_5000\_independent

## [1] 0.12353

# The mean and standard deviations are smaller compared to the simulation with coreelation considered.  
  
# The frequency drops to 12% of the time