Benefit-Cost Analysis

Yinan Zhou

June 2, 2024

Overview and Rationale

In this project I will implement the simulation techniques learned in the module to select the best project among two or more projects under consideration. A primary tool in such selections will be the benefit-cost analysis for each project.

Problem

Corporations must select among many projects that are under consideration by the management. Their primary instrument for evaluating and selecting among the available projects is the benefit-cost analysis. In this analysis, both the annual benefits and the annual costs deriving from a project are estimated in several different categories. Then the total benefit is divided by the total cost to produce a benefit-cost ratio. This ratio is then used by corporations to compare numerous projects under consideration. A benefit-cost ratio greater than 1.0 indicates that the benefits are greater than the costs, and the higher a project's benefit-cost ratio, the more likely it is to be selected over projects with lower ratios. Currently, the JET Corporation is evaluating two dam project constructions, one in southwest Georgia (Dam #1) and the other in North Carolina (Dam #2). The company has identified six areas of benefits: improved navigation, hydroelectric power, fish and wildlife, recreation, flood control, and the commercial development of the area. Furthermore, there are three estimates available for each type benefit – a minimum possible value, a most likely value (i.e., a mode or peak), and a maximum possible value. For the costs, two categories associated with a construction project of this type have been identified: the total capital cost, annualized over 30 years (at a rate specified by the creditors and the government), and the annual operations and maintenance costs. These benefits and costs estimations for both dam projects (in millions of dollars) are as follows:

Table 1: Benefits and costs for the Dam #1 construction project in millions of dollars

Dam #1: Benefits & Costs

	Estimate		
Benefit	*Minimum*	*Mode*	*Maximum*
Improved navigation **B1**	1.1	2	2.8
Hydroelectric power **B2**	8	12	14.9
Fish and wildlife **B3**	1.4	1.4	2.2
Recreation **B4**	6.5	9.8	14.6
Flood control **B5**	1.7	2.4	3.6
Commercial development **B6**	0	1.6	2.4
Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	13.2	14.2	19.1
Operations & Maintenance **C2**	3.5	4.9	7.4

Table 2: Benefits and costs for the Dam #2 construction project in millions of dollars

Dam # 2: Benefits & Costs

	Estimate		
Benefit	*Minimum*	*Mode*	*Maximum*
Improved navigation **B1**	2.1	3	4.8
Hydroelectric power **B2**	8.7	12.2	13.6
Fish and wildlife **B3**	2.3	3	3
Recreation **B4**	5.9	8.7	15
Flood control **B5**	0	3.4	3.4
Commercial development **B6**	0	1.2	1.8
Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	12.8	15.8	20.1
Operations & Maintenance **C2**	3.8	5.7	8

Part 1

(i) Perform a simulation of 10,000 benefit-cost ratios for Dam #1 project and 10,000 such simulations for Dam #2 project. Note that the two simulations should be independent of each other. Let these two ratios be denoted by α_1 and α_2 for the dams 1 and 2 projects respectively.

Simulation Setup

I performed a simulation to compare the benefits and costs of dam 1 and dam 2. It begins by initializing several vectors to store the results of 10,000 iterations of simulated benefits, costs, and benefit-cost ratios for each dam. These vectors include **cost1**, **cost2**, **benifit1**, **benifit2**, **ratio1**, and **ratio2**. The simulation employs a triangular distribution to generate random variates, facilitated by the **getRandomVariate** function, which takes three parameters: *aa* (minimum), *cc* (mode), and *bb* (maximum). This function determines whether to use the ascending or descending portion of the triangular distribution based on a uniform random number and calculates the corresponding random variate.

```
> # Part1
> # (i)
> set.seed(123)
> K<-10000
> cost1<-vector()
> cost2<-vector()
> benifit1<-vector()
> benifit2<-vector()
> ratio1 <-vector()
> ratio2 <-vector()
>
> getRandomVariate <-function(a,c,b){
+   Fc <- (c-a)/(b-a)
+   U <- runif(1)
+   x <- 0
+   if (U<Fc){
+     x<-a+sqrt(U*(b-a)*(c-a))
+   }
+   else{
+     x<-b-sqrt((1-U)*(b-a)*(b-c))
+   }
+   return (x)
+ }
```

Simulation Loop

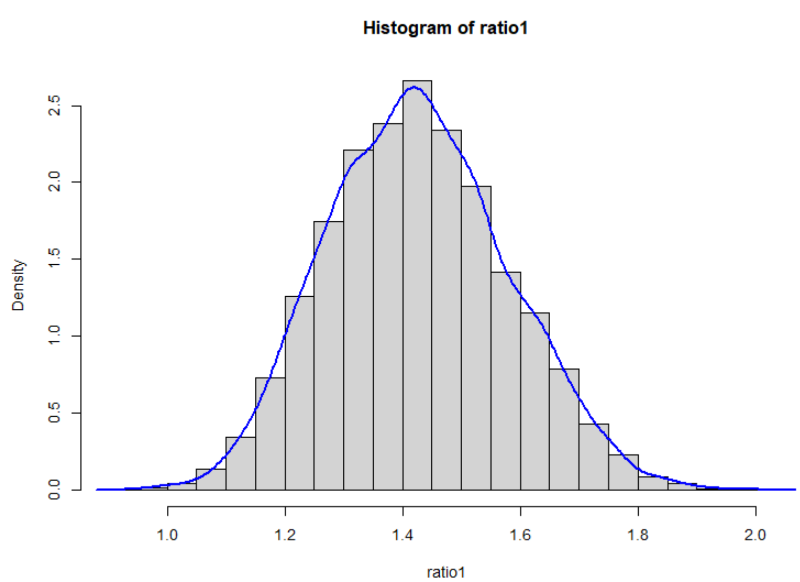
Within a loop that iterates 10,000 times, the code simulates benefits and costs for both dams. For Dam 1, it generates six benefit values (**B1** to **B6**) and two cost values (**C1** and **C2**) using the "**getRandomVariate"** function with predefined parameters. The total benefits and costs are then summed and stored in the respective vectors (**benifit1** and **cost1**). The benefit-cost ratio for each iteration is calculated and stored in the **ratio1** vector. The same process is repeated for Dam 2, with a different set of parameters for the benefits (**B1_d2** to **B6_d2**) and costs (**C1_d2** and **C2_d2**). These values are similarly stored in **benifit2**, **cost2**, and **ratio2**.

```
> for (j in 1:K){
+   #Dam1
+   B1 <- getRandomVariate(1.1, 2, 2.8)
+   B2 <- getRandomVariate(8, 12, 14.9)
+   B3 <- getRandomVariate(1.4, 1.4, 2.2)
+   B4 <- getRandomVariate(6.5, 9.8, 14.6)
+   B5 <- getRandomVariate(1.7, 2.4, 3.6)
+   B6 <- getRandomVariate(0, 1.6, 2.4)
+   C1 <- getRandomVariate(13.2, 14.2, 19.1)
+   C2 <- getRandomVariate(3.5, 4.9, 7.4)
+   benifit1[j] <- B1 + B2 + B3 + B4 + B5 + B6
+   cost1[j] <- C1 + C2
+   ratio1[j] <- benifit1[j]/cost1[j]
+
+   ################################################################################
+   #Dam2
+   B1_d2 <- getRandomVariate(2.1, 3, 4.8)
+   B2_d2 <- getRandomVariate(8.7, 12.2, 13.6)
+   B3_d2 <- getRandomVariate(2.3, 3, 3)
+   B4_d2 <- getRandomVariate(5.9, 8.7, 15)
+   B5_d2 <- getRandomVariate(0, 3.4, 3.4)
+   B6_d2 <- getRandomVariate(0, 1.2, 1.8)
+   C1_d2 <- getRandomVariate(12.8, 15.8, 20.1)
+   C2_d2 <- getRandomVariate(3.8, 5.7, 8)
+   benifit2[j] <- B1_d2 + B2_d2 + B3_d2 + B4_d2 + B5_d2 + B6_d2
+   cost2[j] <- C1_d2 + C2_d2
+   ratio2[j] <- benifit2[j]/cost2[j]
+ }
```

(ii) Construct both a tabular and a graphical frequency distribution for α_1 and α_2 separately (a tabular and a graphical distribution for α_1, and a tabular and a graphical distribution for α_2, a total of 4 distributions). In your report, include only the graphical distributions and comment on the shape of each distribution.

- **Graphical frequency distribution for ratio1(α_1)**

```
> #Dam 1
> hist_ratio1 <- hist(ratio1, breaks =20, freq = F)
> lines(density(ratio1), lwd=2, col="blue")
```



The graphical frequency distribution for **ratio1(α_1)** exhibits a nearly symmetrical bell shape.

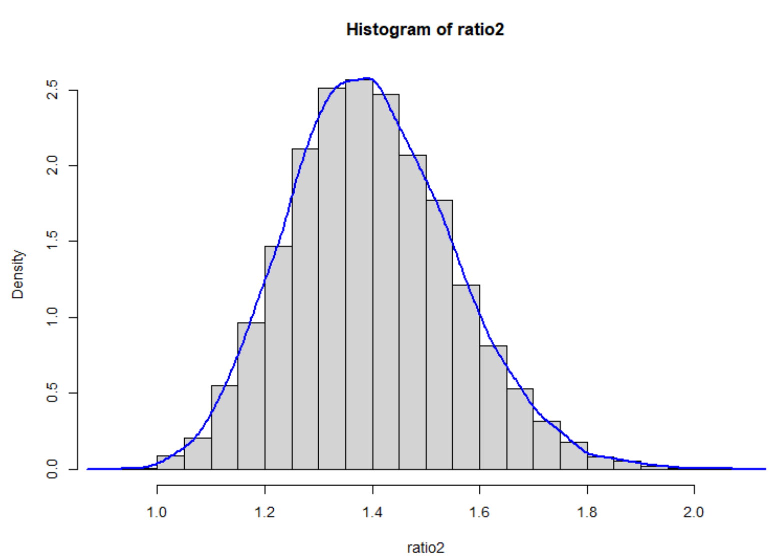
- **Tabular frequency distribution for ratio1(α_1)**

```
> # Create bins and cut the ratio1 data into bins
> breaks <- seq(0.95, 2.10, by = 0.05)
> ratio1_cut <- cut(ratio1, breaks, right = FALSE, include.lowest = TRUE)
> # Create a frequency table
> freq_table1 <- table(ratio1_cut)
> freq_df1 <- as.data.frame(freq_table1)
> colnames(freq_df1) <- c("Ratio1 Range", "Frequency")
```

Ratio1 Range	Frequency
[0.95,1)	8
[1,1.05)	20
[1.05,1.1)	67
[1.1,1.15)	172
[1.15,1.2)	366
[1.2,1.25)	629
[1.25,1.3)	873
[1.3,1.35)	1105
[1.35,1.4)	1192
[1.4,1.45)	1330
[1.45,1.5)	1169
[1.5,1.55)	989
[1.55,1.6)	707
[1.6,1.65)	577
[1.65,1.7)	392
[1.7,1.75)	214
[1.75,1.8)	114
[1.8,1.85)	43
[1.85,1.9)	23
[1.9,1.95)	5
[1.95,2)	2
[2,2.05)	1
[2.05,2.1]	0

- **Graphical frequency distribution for ratio2(α_2)**

```
> #Dam 2
> hist_ratio2 <-hist(ratio2, breaks =20, freq = F)
> lines(density(ratio2), lwd=2, col="blue")
```



The graphical frequency distribution for ratio2 (α2) exhibits a slightly positively skewed bell shape.

- **Tabular frequency distribution for ratio2(α2)**

```
> # Create bins and cut the ratio2 data into bins
> breaks <- seq(0.95, 2.10, by = 0.05)
> ratio2_cut <- cut(ratio2, breaks, right = FALSE, include.lowest = TRUE)
> # Create a frequency table
> freq_table2 <- table(ratio2_cut)
> freq_df2 <- as.data.frame(freq_table2)
> colnames(freq_df2) <- c("Ratio2 Range", "Frequency")
```

Ratio2 Range	Frequency
[0.95,1)	4
[1,1.05)	45
[1.05,1.1)	102
[1.1,1.15)	275
[1.15,1.2)	482
[1.2,1.25)	733
[1.25,1.3)	1054
[1.3,1.35)	1255
[1.35,1.4)	1282
[1.4,1.45)	1235
[1.45,1.5)	1033
[1.5,1.55)	887
[1.55,1.6)	607
[1.6,1.65)	408
[1.65,1.7)	264
[1.7,1.75)	158
[1.75,1.8)	88
[1.8,1.85)	41
[1.85,1.9)	27
[1.9,1.95)	10
[1.95,2)	5
[2,2.05)	3
[2.05,2.1]	1

(iii) For each of the two dam projects, perform the necessary calculations in order to complete the following table. Excel users should create the table in Excel with all cells being occupied by the appropriate formulas, and R users should display the table as a "data frame". Remember to create two such tables – one table for Dam #1 and another table for Dam #2. Include both tables in your report.

I calculated and compared the observed and theoretical values for the benefits, costs, and benefit-cost ratios of Dam 1 and Dam 2.

Firstly, I defined three primary functions to calculate theoretical mean and theoretical standard deviation (SD) for the benefits and costs: **cal_Mean_Var**, **getMean**, and **getVar**.

- The **"cal_Mean_Var"** function calculates the mean and variance for a triangular distribution given its minimum (**a**), mode (**c**), and maximum (**b**). It returns a list containing the mean and variance.

- The **"getMean"** function calculates the total mean by summing up the means of individual components (benefits or costs).
- The **"getVar"** function calculates the total variance by summing up the variances of individual components (benefits or costs).

```
> # (iii)
> cal_Mean_Var <- function(a,c,b){
+   mean = (a+b+c)/3.0
+   var = (a^2+b^2+c^2-a*b-a*c-b*c)/18.0
+   return (list(mean = mean, var = var))
+ }
>
> getMean <- function(B,n){
+   total_mean<-0
+   for (j in 1:n){
+     total_mean<-total_mean+B[[j]]$mean
+   }
+   return (total_mean)
+ }
>
> getVar <- function(B,n){
+   total_var<-0
+   for (j in 1:n){
+     total_var<-total_var+B[[j]]$var
+   }
+   return (total_var)
+ }
```

For Dam 1, the theoretical values for benefits and costs are calculated using the **"cal_Mean_Var"** function for each component of benefits and costs. These values are stored in lists (**"B_m_v"** for benefits and **"C_m_v"** for costs). The total mean and variance for the benefits and costs are then calculated using the **"getMean"** and **"getVar"** functions. The observed values, which include the mean and standard deviation of the benefits, costs, and benefit-cost ratios, are computed from the simulation results and stored in a vector. Then, the theoretical and observed values for Dam 1 are combined into a data frame for easier comparison.

```
> #dam1
> B_m_v <-list()
> B_m_v[[1]] <- cal_Mean_Var(1.1, 2, 2.8)
> B_m_v[[2]] <- cal_Mean_Var(8, 12, 14.9)
> B_m_v[[3]] <- cal_Mean_Var(1.4, 1.4, 2.2)
> B_m_v[[4]] <- cal_Mean_Var(6.5, 9.8, 14.6)
> B_m_v[[5]] <- cal_Mean_Var(1.7, 2.4, 3.6)
> B_m_v[[6]] <- cal_Mean_Var(0, 1.6, 2.4)
> C_m_v <-list()
> C_m_v[[1]] <- cal_Mean_Var(13.2, 14.2, 19.1)
> C_m_v[[2]] <- cal_Mean_Var(3.5, 4.9, 7.4)
>
> observed_values <- c(mean(benifit1), sd(benifit1), mean(cost1), sd(cost1), mean(ratio1), sd(ratio1))
> theoretical_values <- c(getMean(B_m_v,6),sqrt(getVar(B_m_v,6)), getMean(C_m_v,2),sqrt(getVar(C_m_v,2)),NA,NA)
>
> dam1_df <- data.frame(
+   Dam1 = c("Mean of the Total Benefits",
+            "SD of the Total Benefits",
+            "Mean of the Total Cost",
+            "SD of the Total Cost",
+            "Mean of the Benefit-cost Ratio",
+            "SD of the Benefit-cost Ratio"),
+   Observed = observed_values,
+   Theoretical = theoretical_values
+ )
```

The same process was repeated for Dam 2. The theoretical values for benefits and costs were calculated, and the observed values were computed from the simulation results. The theoretical and observed values were then combined into a data frame for Dam 2.

```
> #dam2
> B_m_v <-list()
> B_m_v[[1]] <- cal_Mean_Var(2.1, 3, 4.8)
> B_m_v[[2]] <- cal_Mean_Var(8.7, 12.2, 13.6)
> B_m_v[[3]] <- cal_Mean_Var(2.3, 3, 3)
> B_m_v[[4]] <- cal_Mean_Var(5.9, 8.7, 15)
> B_m_v[[5]] <- cal_Mean_Var(0, 3.4, 3.4)
> B_m_v[[6]] <- cal_Mean_Var(0, 1.2, 1.8)
> C_m_v <-list()
> C_m_v[[1]] <- cal_Mean_Var(12.8, 15.8, 20.1)
> C_m_v[[2]] <- cal_Mean_Var(3.8, 5.7, 8)
>
> observed_values <- c(mean(benifit2), sd(benifit2), mean(cost2), sd(cost2), mean(ratio2), sd(ratio2))
> theoretical_values <- c(getMean(B_m_v,6),sqrt(getVar(B_m_v,6)), getMean(C_m_v,2),sqrt(getVar(C_m_v,2)),NA,NA)
> dam2_df <- data.frame(
+   Dam2 = c("Mean of the Total Benefits",
+            "SD of the Total Benefits",
+            "Mean of the Total Cost",
+            "SD of the Total Cost",
+            "Mean of the Benefit-cost Ratio",
+            "SD of the Benefit-cost Ratio"),
+   Observed = observed_values,
+   Theoretical = theoretical_values
+ )
```

Finally, the data frames for both Dam 1 and Dam 2 are printed as follows:

```
> print(dam1_df)
                            Dam1   Observed Theoretical
1     Mean of the Total Benefits 29.4545509   29.466667
2       SD of the Total Benefits  2.3131798    2.307476
3         Mean of the Total Cost 20.7661672   20.766667
4           SD of the Total Cost  1.5202478    1.520599
5 Mean of the Benefit-cost Ratio  1.4258994          NA
6   SD of the Benefit-cost Ratio  0.1524012          NA
> print(dam2_df)
                            Dam2   Observed Theoretical
1     Mean of the Total Benefits 30.7036845   30.700000
2       SD of the Total Benefits  2.3805585    2.409703
3         Mean of the Total Cost 22.0784357   22.066667
4           SD of the Total Cost  1.7270007    1.726589
5 Mean of the Benefit-cost Ratio  1.3991265          NA
6   SD of the Benefit-cost Ratio  0.1538984          NA
```

Part 2

Use your observation in Question (ii) of Part 1 to select a theoretical probability distribution that, in your judgement, is a good fit for the distribution of α_1. Next, use the Chi-squared Goodness-of-fit test to verify whether your selected distribution was a good fit for the distribution of α_1. Describe the rational for your choice of the probability distribution and a description of the outcomes of your Chi-squared test in your report. In particular, indicate the values of the Chi-squared test statistic and the P-value of your test in your report, and interpret those values.

Initial Analysis

First, I calculated the skewness and kurtosis of the benefit-cost ratios for both Dam 1 and Dam 2 as follows:

```
> # Part2
> # install.packages("fBasics")
> library(fBasics)
> skewness(ratio1)
[1] 0.1700601
attr(,"method")
[1] "moment"
> kurtosis(ratio1)
[1] -0.19364
attr(,"method")
[1] "excess"
> skewness(ratio2)
[1] 0.3224457
attr(,"method")
[1] "moment"
> kurtosis(ratio2)
[1] 0.09235072
attr(,"method")
[1] "excess"
```

From the outcome R returned, the slight positive skewness for both **ratio1** and **ratio2** suggests that both distributions have longer tails on the right side, with **ratio2** being slightly more skewed than **ratio1**. The kurtosis values indicate that **ratio1** has a distribution with slightly lighter tails and a flatter peak than the normal distribution, while **ratio2** has a distribution that is very similar to the normal distribution in terms of tail heaviness and peak sharpness.

Distribution Fitting

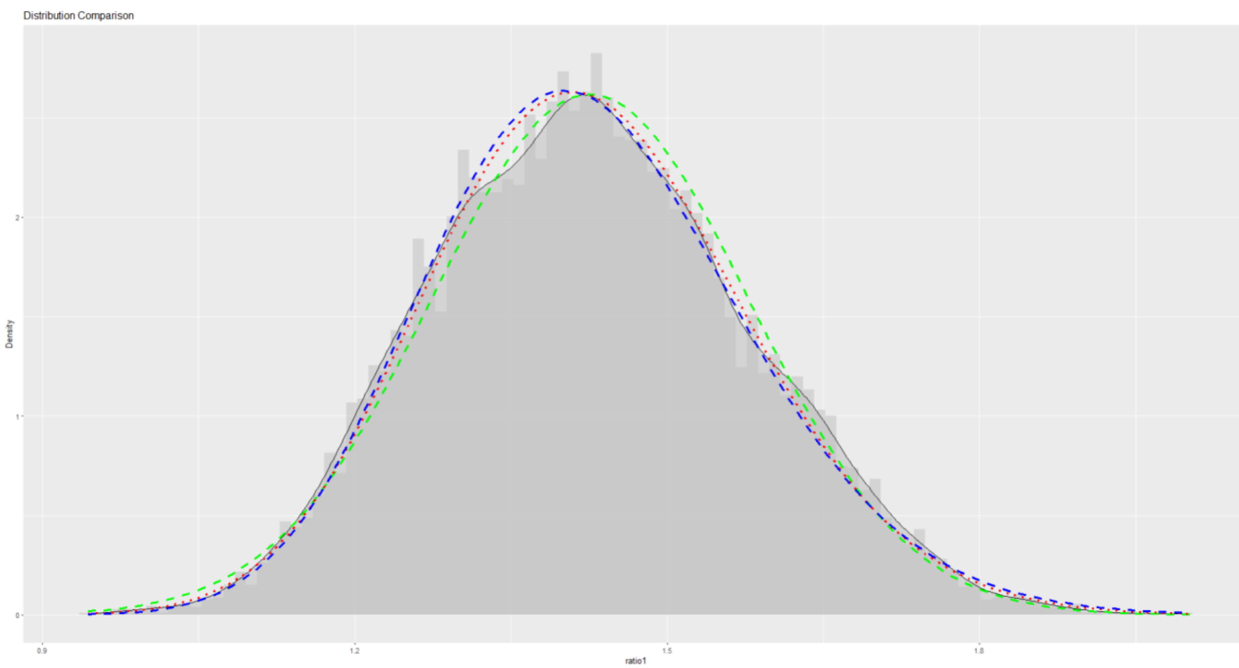
Next, I fitted three different theoretical distributions to the **ratio1** data: Normal, Gamma, and Log-normal distributions. This step estimates the parameters for each distribution based on the observed data.

```
> fit_normal<-fitdistr(ratio1,"normal")
> fit_gamma<-fitdistr(ratio1,"gamma")
> fit_lognormal<-fitdistr(ratio1,"log-normal")
> fit_normal<-fitdistr(ratio1,"normal")
> fit_gamma<-fitdistr(ratio1,"gamma")
> fit_lognormal<-fitdistr(ratio1,"log-normal")
> fit_normal
      mean           sd
  1.425899422   0.152393628
 (0.001523936) (0.001077586)
> fit_gamma
     shape         rate
  87.5388294   61.3920103
 ( 1.2355643) ( 0.8689958)
> fit_lognormal
     meanlog        sdlog
  0.3490701033   0.1073093236
 (0.0010730932) (0.0007587915)
```

Visual Comparison

Then, I created a histogram and density plot of the **ratio1** data and overlays the probability density functions of the fitted distributions. This visual comparison helps to assess how well each distribution fits the observed data for α_1.

```
> df <- data.frame(ratio1)
> p <- ggplot(df, aes(x=ratio1)) +
+   geom_histogram(aes(y=..density..), bins=100, fill="gray", alpha=0.5) +
+   geom_density(aes(y=..density..), fill="gray", alpha=0.5) +
+   ggtitle("Distribution Comparison") +
+   xlab("ratio1") +
+   ylab("Density")
>
> p <- p + stat_function(fun=dnorm, args=list(mean=fit_normal$estimate["mean"], sd=fit_normal$estimate["sd"]),
+                        color="green", size=1.2, linetype="dashed")
>
> p <- p + stat_function(fun=dgamma, args=list(shape=fit_gamma$estimate["shape"], rate=fit_gamma$estimate["rate"]),
+                        color="red", size=1.2, linetype="dotted")
>
> p <- p + stat_function(fun=dlnorm, args=list(meanlog=fit_lognormal$estimate["meanlog"], sdlog=fit_lognormal$estimate["sdlog"]),
+                        color="blue", size=1.2, linetype="dashed")
> p
```



From the plot, we can find that the red dotted line fits the observed data for α_1 best, which is the **gamma distribution**.

Chi-squared Goodness-of-Fit Test

To quantitatively verify the goodness of fit, I performed the Chi-squared Goodness-of-Fit test for the gamma distribution. The Chi-squared test compares the observed frequencies in each bin of the histogram to the expected frequencies under gamma distribution, calculating the Chi-squared test statistic and p-value.

```
> hist_ratio1 <- hist(ratio1, breaks =180, freq = F)
> observed_counts <- hist_ratio1$counts
> bin_edges <- hist_ratio1$breaks
> # Calculate expected counts for gamma distribution
> expected_gamma <- diff(pgamma(bin_edges, shape=fit_gamma$estimate["shape"], rate=fit_gamma$estimate["rate"])) * length(ratio1)
> # Perform Chi-squared Goodness-of-Fit tests
> chi_square_gamma <- chisq.test(observed_counts, p=expected_gamma, rescale.p = TRUE)
```

```
> chi_square_gamma

        Chi-squared test for given probabilities

data:  observed_counts
X-squared = 210.35, df = 212, p-value = 0.5191
```

The X-squared (Chi-squared) value is a statistic that measures the discrepancy between the observed and expected frequencies in the data. In this case, it is 210.35.

The p-value associated with the Chi-squared test is 0.5191. This p-value is compared to a significance level (commonly 0.05) to determine the statistical significance of the Chi-squared test result.

In this case, the p-value of 0.5191 is well above the commonly used significance level of 0.05. This suggests that there is no significant difference between the observed data and the gamma distribution (theoretical distribution) used in the Chi-squared test. Therefore, we fail to reject the null hypothesis, indicating that the gamma distribution is a good fit for the observed data.

Part 3

(i) Use the results of your simulations and perform the necessary calculations in order to complete the table below. Excel users should create the table in Excel with all cells being occupied by the appropriate formulas, and R users should display the table as a "data frame". Include the completed table in your report.

Firstly, a custom function named **"getRatioP (ratio, alpha)"** was defined. This function calculates the probability that a value in the given **ratio** dataset is greater than a specified threshold **alpha**. Then, I calculated probabilities for both **ratio1** and **ratio2** using different threshold values, such as 2, 1.8, 1.5, 1.2, and 1.

```
> # Part3
> # (i)
> getRatioP <- function(ratio,alpha){
+   P = sum(ratio>alpha)/length(ratio)
+   return (P)
+ }
> P1 <- getRatioP(ratio1,2)
> P2 <- getRatioP(ratio1,1.8)
> P3 <- getRatioP(ratio1,1.5)
> P4 <- getRatioP(ratio1,1.2)
> P5 <- getRatioP(ratio1,1)
>
> P6 <- getRatioP(ratio2,2)
> P7 <- getRatioP(ratio2,1.8)
> P8 <- getRatioP(ratio2,1.5)
> P9 <- getRatioP(ratio2,1.2)
> P10 <- getRatioP(ratio2,1)
```

After computing probabilities, I proceeded to calculate various summary statistics and metrics for both **ratio1** and **ratio2**. These metrics include minimum, maximum, mean, median, variance, standard deviation, and skewness.

```
> dam1 <-c(min(ratio1), max(ratio1), mean(ratio1), median(ratio1), var(ratio1), sd(ratio1), skewness(ratio1), P1, P2, P3, P4, P5)
> dam2 <-c(min(ratio2), max(ratio2), mean(ratio2), median(ratio2), var(ratio2), sd(ratio2), skewness(ratio2), P6, P7, P8, P9, P10)
```

All the calculated metrics, probabilities, and summary statistics are then organized into a structured data frame named **"ratio_comparison"** as follows:

```
> ratio_comparison <- data.frame(
+   Metric = c("Minimum",
+              "Maximum",
+              "Mean",
+              "Median",
+              "Variance",
+              "Standard Deviation",
+              "Skewness",
+              "P(alpha>2)",
+              "P(alpha>1.8)",
+              "P(alpha>1.5)",
+              "P(alpha>1.2)",
+              "P(alpha>1)"),
+   ratio1 = dam1,
+   ratio2 = dam2
+ )
> ratio_comparison
               Metric     ratio1     ratio2
1             Minimum 0.94383439 0.93531057
2             Maximum 2.00218864 2.06548358
3                Mean 1.42589942 1.39912646
4              Median 1.42171882 1.39174508
5            Variance 0.02322614 0.02368471
6  Standard Deviation 0.15240125 0.15389837
7            Skewness 0.17006011 0.32244572
8          P(alpha>2) 0.00010000 0.00040000
9        P(alpha>1.8) 0.00740000 0.00870000
10       P(alpha>1.5) 0.30670000 0.24990000
11       P(alpha>1.2) 0.93650000 0.90910000
12         P(alpha>1) 0.99900000 0.99950000
```

(ii) In your report, use your observations of the results obtained in parts 1-3 to recommend one of two projects to the management. Explain all your rationales for the project that you have recommended. In particular, include with the final conclusion of your report an estimate for the probability that α_1 will be greater than α_2.

I calculated the probability that a value from **ratio1** is greater than a corresponding value from **ratio2**. From the result, the estimated probability that Dam 1's benefit-cost ratio will be greater than Dam 2's is approximately 55.16%.

```
> # (ii)
> #probability of ratio1>ratio2
> sum(ratio1>ratio2)/length(ratio1)
[1] 0.5516
```

Based on the comparison between the benefit-cost ratios of Dam 1 (**ratio1**) and Dam 2 (**ratio2**) as shown in **"ratio_comparison"** and the calculated probability that Dam 1's ratio will be greater than Dam 2's, here are the rationales for the recommended project:

1. **Minimum and Maximum Values:**
   Ratio1 has a slightly higher minimum and maximum value compared to **ratio2**, indicating potentially better performance in extreme scenarios.

2. **Mean and Median:**
   Ratio1 also shows a slightly higher mean and median, suggesting a central tendency towards better outcomes on average.
3. **Variance and Standard Deviation:**
   Ratio1 also shows a slightly lower variance and standard deviation, indicating lower levels of variability in outcomes.
4. **Skewness:**
   Ratio1 has lower skewness compared to ratio2, implying a more symmetric distribution and potentially more stable outcomes.
5. **Probabilities (P(alpha>2) to P(alpha>1)):**
   The probabilities of **ratio1** being greater than specific thresholds (2, 1.8, 1.5, 1.2, and 1) are generally higher than those of ratio2, indicating a higher likelihood of favorable outcomes for ratio1.
6. **Probability (ratio 1 > ratio 2):**
   The estimated probability that Dam 1's benefit-cost ratio will be greater than Dam 2's is approximately 55.16%, further supporting Dam 1 as the preferred option.

Overall, I recommend **Dam 1** due to its higher probabilities of success, better central tendency, and potential stability in outcomes.