Logo

Description automatically generated

**College of Professional Studies**

**Northeastern University San Jose**

**MPS Analytics**

**Course: ALY6015: Introduction to Enterprise Analytics**

**Assignment:**

Module 2 Project-  Benefit-Cost Analysis

**Submitted on:**

March 5, 2023

**Submitted to:**  **Submitted by:**

Prof: AZADEH MOBASHER NIKSHITA RANGANATHAN

# **ABSTRACT**

The computing technique of Monte Carlo simulation simulates complex processes and systems using random sampling. Scientists working on the Manhattan Project during World War II created the Monte Carlo simulation technique to simulate neutron activity in nuclear reactors. It is named after the famous casino in Monte Carlo, which is known for its games of chance. The technique is widely used in various fields, including finance, engineering, physics, and scientific research.

The probability distributions of the input variables are used to create a large number of random simulations throughout a Monte Carlo simulation. The outcomes of each simulation are then combined to create a distribution of possible outcomes that can be used to compute probabilities and draw conclusions.

Here are some uses of Monte Carlo simulation in enterprises:

* It can be used to conduct sensitivity analysis, which involves assessing how changes in input variables affect the output. This way companies can recognize critical input variables.
* It is an effective method for optimizing systems that have complex and nonlinear objective functions, as well as constraints. By generating multiple simulations and evaluating the results, it can help identify the optimal value of the objective function, taking into account the constraints placed on the system.
* It is a valuable tool in project management that can be used to estimate project duration and identify potential risks. It can be useful for businesses to identify critical activities and allocate resources effectively to ensure project success.
* Monte Carlo simulation can be used in supply chain management to model inventory and demand uncertainties. It can benefit businesses to optimize inventory levels and improve supply chain efficiency.

The triangular distribution is a continuous probability distribution that is often used in statistics to model uncertain variables with a limited range of possible values. It is called a triangular distribution because its probability density function looks like a triangle, with the peak located at the mode of the distribution.

One of the advantages of using the triangular distribution is its simplicity and flexibility. It is relatively easy to estimate the three parameters of the distribution based on available data or expert opinions.

Some examples of real-world applications of the triangular distribution include:

* Estimating the time required to complete a task in project management
* Estimating the demand for a new product in marketing
* Modeling the uncertainty in financial forecasts
* Simulating the variation in arrival times of customers at a service facility

**INTRODUCTION**

This project involves conducting a Benefit-Cost Analysis for projects available to a corporation, where estimates are made in various categories for the annualized benefits and costs of the projects. The Benefit-Cost Ratio of a project is obtained by dividing the total benefits by the total costs. We have to note that the higher the benefit-cost ratio, the more likely the project will be selected.

The problem at hand involves JET Corporation's evaluation of two dam projects in southwest Georgia and North Carolina, considering benefits in six areas, such as developed route, fish and wildlife, hydroelectric power, flood control, entertainment, and business development. There are three figures for each category: a minimum, a most likely value, and a maximum value. The two cost categories associated with this type of project are the total capital cost (annualized) and the annual operations and maintenance costs.

Through simulation techniques, we will be able to simulate different scenarios and evaluate the benefit-cost ratios of each dam project to assist the organization in making an informed decision. This assignment will provide a practical application of simulation techniques in project selection, a critical task for many organizations.

Below is the data given in the question

Table

Description automatically generated

**Figure 1- Table with benefits and costs (Dam -1)**

**Table

Description automatically generated**

**Figure 2- Table with benefits and costs (Dam -2)**

**ANALYSIS & INTERPRETATION**

**Part-1**

1. **Simulations for 10,000 Benefit-Cost Ratios of Dams 1 and 2**

**Table

Description automatically generated**

**Figure 3- Table with calculations (Dam -1)**

**Table

Description automatically generated**

**Figure 4- Table with calculations (Dam -2)**

* The above tables depict the calculations for triangular distribution for both dams. The theoretical mean and variance is also be calculated based on its three parameters.
* Triangular distribution is defined by three parameters: a **minimum value (a), a maximum value (b), and a mode value (c)**. Together, "a", "b", and "c" determine the shape of the triangular distribution, including its skewness and kurtosis.
* The theoretical mean of a triangular distribution is calculated as the average of its minimum, maximum, and mode values: **E(X) = (a+b+c)/3.**
* The theoretical variance of a triangular distribution is calculated as **(a^2+b^2+c^2-ab-bc-ca)/18.**

**A picture containing text

Description automatically generated**

**Figure 5- Formula for random number generation in a triangular distribution**

The formula for Random number generation uses the cumulative probability distribution function of the triangular distribution to transform a uniform random variable into a triangular random variable.

From a uniform distribution between 0 and 1, we first produce a random value, r. we utilized the built-in function **RAND()** in Excel to generate random values that were then used as inputs.

In the formula –

* K is the ratio of the distance from the minimum to the mode and the distance from the minimum to the maximum, **(K = (c-a) / (b-a))**
* M is the product of the distances from the minimum to the mode and from the mode to the maximum. **(M =(b-a) (c-a))**
* N is the product of the distances from the mode to the maximum and from the minimum to the mode. **(N =(b-a) (b-c))**

**Calendar

Description automatically generated with medium confidence**

**Figure 6- Total Benefits and Total costs**

* There are **six benefit variables**, labeled as B1 through B6, and six corresponding random variables labeled as r\_B1 through r\_B6.
* The same is true for the **two cost variables**, labeled as C1 and C2, and their corresponding random variables r\_C1 and r\_C2.
* The "**Total Benefits**" and "**Total Costs**" columns refer to the sum of all simulated values that are classified as benefits and costs, respectively.

**Table

Description automatically generated**

**Figure 7- Benefit-cost ratios for Dams 1 and 2**

A higher benefit-cost ratio is generally seen as desirable and is often used as a criterion for evaluating the feasibility and desirability of a proposed project.

1. **Tabular and a graphical frequency distribution**

**DAM-1**

**Table

Description automatically generated**

**Figure 8- Minimum, Maximum, Range, Bins, and Class width for Dam 1**

In this case, the distribution has been divided into **100** bins of equal width **(0.01)** between **1.01** and **2.00**, with a total count of **10,000** observations.

**Table

Description automatically generated**

**Figure 9- Tabular frequency distribution for Dam 1**

Using this table, we can summarize the distribution of the data and make observations about the spread and shape of the data. The average of the class left and class right limits is used to determine each class's midpoint. Class frequency represents the number of times a data value falls within the given range of values for each interval.

**Chart, histogram

Description automatically generated**

**Figure 10- Graphical frequency distribution for Dam 1**

Based on the shape of the histogram, it may be possible to make an initial guess about the distribution of the data. Histogram for Dam 1 suggests that the distribution of the Benefit-Cost ratios could be either **normal** or **gamma distribution**.

A normal distribution is a bell-shaped curve, with the mean, median, and mode being equal and the data being symmetric around the mean and gamma distribution is a positively skewed distribution with a long tail on the right side, and it is often used to model data that represents waiting times or durations.

**Table

Description automatically generated**

**Figure 11- Minimum, Maximum, Range, Bins, and Class width for Dam 2**

For Dam 2, we have a minimum value of **0.93**, a maximum value of **2.01**, and a range of **1.08**. The distribution contains **10,000** observations, which are grouped into **100** classes or bins with a class width of **0.01**.

Table

Description automatically generated

**Figure 12- Tabular frequency distribution for Dam 2**

Chart, histogram

Description automatically generated

**Figure 13- Graphical frequency distribution for Dam 2**

The Benefit-Cost ratios of Dam 2 also appear to be distributed according to either a **normal** or **gamma distribution**.

1. **For the two dam projects, perform the necessary calculations in order to complete the table**

Table

Description automatically generated

**Figure 14- Additional Calculations for Dam 1 and 2**

We can understand summary statistics for Dam 1 and Dam 2, including observed and theoretical values for various measures in these tables.

For Dam 1, the mean of total benefits is slightly higher than the theoretical value, and the mean of the total cost is less than the theoretical value. Both the benefit and cost standard deviations are considerably less than their theoretical estimates. The benefit-cost ratio for Dam 1 has an observed mean of **1.446** and an observed standard deviation of **0.15**.

For Dam 2, the mean of total benefits is greater than the theoretical value, while the mean of the total cost is also slightly more than the theoretical value. The standard deviations for both benefits and costs are very close to their theoretical values. The benefit-cost ratio for Dam 2 has an observed mean of **1.399** and an observed standard deviation of **0.155**.

**Part-2**

**Chi-squared Goodness-of-fit test**

A graphical representation alone is not sufficient to determine the distribution of the data. The chi-squared goodness of fit test can help us determine whether the observed data follows a particular distribution or not, and if so, which distribution it follows. I conducted two chi-square tests to find the distribution for benefit-cost ratio of Dam 1, first will be triangular distribution.

Table

Description automatically generated

**Figure 15- Calculations for the Triangular distribution (Benefit Cost ratio Dam 1)**

* **Null hypothesis:** The observed data follows a triangular distribution with parameters a (minimum value), b (maximum value), and c (mode).
* **Alternative hypothesis:** The observed data does not follow a triangular distribution with parameters a(minimum value), b (maximum value), and c (mode).

A picture containing graphical user interface

Description automatically generated

**Figure 16- Theoretical probability, Expected frequency for Triangular distribution (Benefit Cost ratio Dam 1)**

The theoretical probability for a triangular distribution is calculated using the following formulas:

**If x ≤ c, then P(X ≤ x) = (1/A) \* (x - a)^2**

**If x > c, then P(X ≤ x) = 1 - (1/B) \* (b - x)^2**

Table

Description automatically generated

**Figure 17- Chi square test, p value for Triangular distribution (Benefit Cost ratio Dam 1)**

* The degrees of freedom for the chi-squared test depends on the number of bins or intervals used in the histogram and the number of parameters in the theoretical distribution being tested. In this case, the number of bins is 100 and there are 3 parameters so hence **degree of freedom** is **96**.
* We can use the **CHISQ.DIST()** function with the series of theoretical and observed frequencies, assuming an **alpha value of 0.05** which corresponds to a 95% confidence level.
* The P-value is **0.000**, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis at the 95% confidence level.
* It is important to note that the P-value may vary with different random values or iterations.
* Therefore, we can conclude that there is sufficient evidence to suggest that the observed data significantly deviates from a triangular distribution with the given parameters. In other words, **the data does not follow triangular distribution**.

Next, we will be checking if the distribution follows the gamma distribution.

Table

Description automatically generated

**Figure 18- Mean, Variance, Alpha, and Beta values for Gamma distribution (Benefit Cost ratio Dam 1)**

* **Null hypothesis:** The observed data follows a gamma distribution with parameters alpha and beta.
* **Alternative hypothesis:** The observed data does not follow a gamma distribution with parameters alpha and beta.
* **Alpha = Mean^2 / Variance**
* **Beta = Variance / Mean**

Table

Description automatically generated

**Figure 19- Theoretical probability, Expected frequency for Gamma distribution (Benefit Cost ratio Dam 1)**

* We can see the observed frequency, theoretical probability, expected frequency, and the chi-squared test statistic for each class interval of the Benefit-Cost ratios of Dam 1 in Figure 19.
* The **GAMMA.DIST** function in Excel is used to calculate the theoretical probability of a gamma distribution given the parameters of the distribution, such as the shape parameter (alpha) and the rate parameter (beta).

Table

Description automatically generated

**Figure 20- Chi-square test, P value for Gamma distribution (Benefit Cost ratio Dam 1)**

* For gamma distribution, again the number of bins is 100 but number of parameters is **2** (alpha and beta) so the degree of freedom is **97**.
* **Alpha value is 0.05**
* The P-value is **0.478**, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis at the 95% confidence level.
* It is important to note that the P-value may vary with different random values or iterations.
* Therefore, we can conclude that there is not enough evidence to suggest that the observed data significantly deviates from a gamma distribution with the given parameters. In other words, **the data follows a gamma distribution**.

**Part-3**

1. **Use the results of simulations and perform the necessary calculations in order to complete the table**

**Table

Description automatically generated**

**Figure 21- Benefit Cost ratios for Dam 1 and 2**

**Table

Description automatically generated**

**Figure 22- Statistical Summary (Benefit Cost ratios Dam 1 and 2)**

* Both ratios have similar minimum and maximum values, but α1 has a slightly higher mean and median compared to α2.
* The variance and standard deviation of α2 are slightly higher than α1. This suggests that the observations in α2 are more spread out compared to those in α1.
* The skewness for α2 is higher than α1, indicating that the observations in α2 are more positively skewed than α1.
* Finally, the probability of α1 being greater than α2 is **0.585**

1. **Observations of the results**

It is recommended that management should choose Dam 1 in the Southwest Georgia area instead of Dam 2 in North Carolina. This is based on the following observations of analysis conducted:

* The mean benefit-cost ratio of Dam 1 **(1.441)** is higher than that of Dam 2 **(1.398)**, indicating that the former project is likely to be more profitable.
* The variance of the benefit-cost ratio for Dam 1 **(0.022)** is less than Dam 2 **(0.024)**. This is important because a smaller variance implies less uncertainty and greater predictability of the outcome.
* The probability that 𝛼1 will be greater than 𝛼2 is **0.585** so there is a higher chance that Dam 1 will generate more economic benefits than Dam 2.

**CONCLUSION**

In this Benefit-Cost Analysis, we evaluated two construction projects (Dam 1 and Dam 2) available to a corporation.

We used the Triangular distribution to calculate the theoretical and observed values and performed Monte Carlo simulations to retrieve different outcomes of all the benefits and costs of both projects.

Furthermore, the application of the Chi-Squared Goodness-of-fit test helped to confirm the type of distribution that best fit the data for Benefit-cost ratio (𝛼1) of Dam #1. We found that the distribution follows Gamma Distribution.

Overall, our analysis suggests that the JET corporation should choose the Dam 1 project located in Southwest Georgia over the Dam 2 project located in North Carolina because of the higher Benefit-Cost Ratio it provides to the company and greater P-values in most cases.

**REFERENCES**

Bluman, A. G. (2018). Elementary Statistics, 10th ed. McGraw Hill.

Kabacoff, R. I. (2011). R in action: Data analysis and graphics with R. Manning Publications Co.

Home - RDocumentation. (n.d.). <https://www.rdocumentation.org/>

Gamma Distribution. (2023, February 6). Real Statistics Using Excel.

https://real-statistics.com/other-key-distributions/gamma-distribution/