**Module-2:** **Benefit-Cost Analysis of Dam Construction Projects**

*By: Nithin Reddy.P (NEU ID:002896440)*

*Subject:ALY6050: Introduction to Enterprise Analytics*

*Under the guidance of Shahram Sattar*

*Submission Date:03-Mar-2024*

# **Introduction:**

In this assignment, we are about to conduct cost benefit analysis of two dams (one in Southwest Georgia and other in North Carolina) and recommend one of the two projects to JET Corp. The associated benefits and cost for 30 years annualized period of respective dams are provided in millions, they identified six areas of benefits and two categories associated with cost.

# **Problem:**

The aim is to perform cost benefit analysis, which is obtained by calculating benefit to cost ratio. If its greater then 1, it indicates that there are more benefits and less benefits in the other case. The case study is segregated into three parts. In the initial part, we calculate the profitability ratio of two projects, then conduct frequency distribution and statistics for observed and theoretical values. Secondly, we perform Chi-Square goodness of fit to determine whether there triangular distribution represents uncertainty in the costs and benefits. Finally, we compare findings from both projects and recommend the best project idea.

# **Part-1:**

# **(i): Perform 10,000 simulations on Dam1 and Dam2, calculate the Benefit to cost ratio and display then as variable Alpha1 and Alpha2.**

The given data, for Benefits and Costs for Dam1 and Dam2 have been stored in a matrix for further analysis, please refer to Fig-1. In order to calculate the Total Benefit Cost Ratio for each Dam, I have created a Cost function (Fig-2), it simulates the ratio for mentioned number of iterations. The two arguments **benefits\_dam** and **costs\_dam** store Minimum, Maximum and Mode respectively. The **IF** loop conducts the simulation for 10,000 times and creates random sales for the mentioned arguments. It then finds the Total Benefits and Total Costs for each instance and lastly computes the requested ratio and stores the variables in a vector of benefit-to-cost ratios (**sampels**), the total benefits (**total\_benefits**), and the total costs (**total\_cost**).

**Cost Benefit Analysis:**

As we now have the required data, we have conducted the analysis of cost-benefit and assigned to respective alphas (Fig-3). Moreover, the respective max, min and mode values have been extracted, by unlisten alpha values. To find the mode, we have created a mode\_function (Fig-4 and 5).

**(ii): Tabular and a graphical frequency distribution for alpha1 and 2**

Further by using Table and breaks (40), we have created tables for respective alphas (Fig-6). Furthermore, by using Hist function we have created Frequency distribution of alpha-1 and alpha-2 (Fig-7 and 8). Form the figures, we can see that the distribution follows a Triangle with the minimum, maximum and mean values. Using this assumption that they follow this distribution, we move forward in our analysis to find the theoretical and observed values.

Form Fig-9, we could observe that the mean and standard deviation have been assigned to the respective variables. To find the theoretical values, we have created a function (Fig-10) and the values have been allocated to the respective variables and the values are stored in data frame (Fig-11). The similar approach has been implemented to find the observed and theoretical values of Dam-2 (Fig-12).

# **Part-2: Triangle Distribution and Chi Square Goodness of Fit**

To populate the triangle distribution plot, we have extracted the respective min, max and mode values earlier (Fig-5). We built the pot by creating a **dtriangular** function (Fig-13), by using the earlier derived values (Fig-5), we have built a triangle distribution plot (Fig-14) for alpha-1. From the plot, we could clearly see that the mean, minimum and mode are placed to the left of the peak.

1. **Mean to the Left of the Peak**: In a symmetric triangular distribution (where the peak is at the midpoint between the minimum and maximum), the mean coincides with the peak. However, if the mean is shifted to the left of the peak, it suggests that the distribution is skewed towards the lower end of the range. This skewness indicates that there is a higher probability of observing values below the mean compared to values above the mean.
2. **Minimum to the Left of the Peak**: Similarly, if the minimum value is positioned to the left of the peak, it means that the distribution extends further towards the lower end, and there is a non-zero probability of observing values below the minimum.
3. **Mode to the Left of the Peak**: The mode being to the left of the peak indicates that the most likely value is also skewed towards the lower end of the distribution. This could happen when the underlying phenomenon has a bias or preference towards lower values.

**Chi-Square Hypothesis Test:**

To explore whether triangular distribution adequately represents the uncertainty in the costs and benefits we have conducted goodness of fit test with the below hypothesis.

**H0:** The use of a triangular distribution adequately represents the uncertainty in the costs and benefits associated with cost-benefit analysis.

**H1:** The use of a triangular distribution does not adequately represents the uncertainty in the costs and benefits associated with cost-benefit analysis.

From the test (Fig-15) we could clearly see that the p-value (0.00000000000000022) is less than 0.05, hence we reject the null hypothesis, suggesting that the triangular distribution does not adequately represent the uncertainty in the costs and benefits associated with cost-benefit analysis.

Similarly, the triangular plot and chi-square text is conducted on Dam-2 and we have observed the similar graph and same results as of Dam-1 (Fig-16 and Fig-17).

# **Part-3: Results Simulation**

In the final step, we have used **moments** library and found the skewness of alpha-1 &2 along with the respective min, max, mean, median, variance and respective probabilities being greater than certain values (2, 1.8, 1.5, 1.2, and 1) using mean function (Fig-18). We tried to check whether probability of C-B ratio of Dam 1 will be higher and observed that Dam-1 outperform Dam-2 (Fig-19).

# **Summary:**

Triangle Plot Analysis: A triangle plot was constructed, where the mean, minimum, and mode were observed to be positioned to the left side of the peak. This positioning suggests a skewness towards lower values in the distribution.

Hypothesis Testing: The p-value obtained from a chi-squared goodness-of-fit test for both Dam-1 and Dam-2 was less than 0.05. Therefore, the null hypothesis was rejected, indicating that the use of a triangular distribution does not adequately represent the uncertainty in the costs and benefits associated with the cost-benefit analysis for both dams.

Descriptive Statistics: Descriptive statistics were computed for both Dam-1 (Alpha1) and Dam-2 (Alpha2) projects, including minimum, maximum, mean, median, variance, standard deviation, skewness, and probabilities of alpha values exceeding certain thresholds.

Comparison of Projects: It was estimated that Alpha1 consistently outperforms Alpha2 in terms of their benefits and costs, with a probability of approximately 0.6002.

# **Conclusion:**

In conclusion, it can be concluded that despite the skewness observed in the triangle plot and the rejection of the null hypothesis in the chi-squared test, Alpha1 consistently outperforms Alpha2 across various metrics. This suggests that, despite the limitations of the triangular distribution model, Alpha1 appears to be a more favourable option compared to Alpha2 in terms of its benefits and costs.

**Appendix:**

## **Scenario-1**

### **Fig-1: Given Data**

A screenshot of a computer code

Description automatically generated

### Fig-2: Function for calculating Cost

**A computer screen shot of a program

Description automatically generated**

### Fig-3: Cost and Benefit analysis:

**A close-up of numbers

Description automatically generated**

### A close-up of a number Description automatically generated

### Fig-4: Function for Mode :

A computer screen shot of a function

Description automatically generated

### Fig-5: Extracting Min, Max, Median and Mode

A screenshot of a computer code

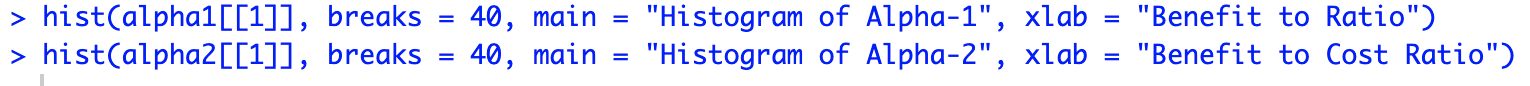
Description automatically generated A screenshot of a computer code

Description automatically generatedFig-6: The Alpha tables

A screenshot of a computer

Description automatically generated

### Fig-7: Frequency distribution



### Fig-8: Histogram of Alpha-1 and Alpha-2:

A graph of a graph

Description automatically generated A graph of a graph

Description automatically generated

### Fig-9: Observed frequencies of Dam-1

A computer code with blue text

Description automatically generated

### Fig-10: Theoretical frequencies of Dam-1

A computer screen shot of a program

Description automatically generated

### Fig-11: Dam-1 Data Frame

A screenshot of a computer

Description automatically generated

### Fig-12: Data Frame for Dam-2

A computer screen shot of a program

Description automatically generated

### Fig-13: dtriangular Function

A computer code with text

Description automatically generated

Fig-14: Dam-1 Plot:

A graph with a line

Description automatically generated

A screenshot of a computer code

Description automatically generated

Fig-15: Dam-1- Chi-Square

A screenshot of a computer code

Description automatically generated

Fig-16: Dam-2 Plot:

A graph with a line

Description automatically generated

Fig-17: Dam-2- Chi-Square

A screenshot of a computer code

Description automatically generated

Fig-18: Results Table

A screenshot of a table

Description automatically generated

A screenshot of a computer program

Description automatically generated

Fig-19: Probability that the C-B ratio of Dam 1 will be higher

A screenshot of a computer code

Description automatically generated

# **References:**

ALY 6050: Module 2 Lab 2 — The Triangular Probability Distributions & Random Number Generation

JIM (22 September 2022): Triangular Distribution in R; [online]: [Triangular Distribution in R | R-bloggers](https://www.r-bloggers.com/2022/09/triangular-distribution-in-r/#google_vignette)

Steven P. P. Sanderson: Understanding the Triangular Distribution and Its Application in R; [online]: [Understanding the Triangular Distribution and Its Application in R | R-bloggers](https://www.r-bloggers.com/2023/10/understanding-the-triangular-distribution-and-its-application-in-r/#google_vignette)

R Document: The Triangular Distribution; [online]: [R: The Triangular Distribution (r-project.org)](https://search.r-project.org/CRAN/refmans/EnvStats/html/Triangular.html)

Kausthub Balaji: Triangular Distribution; [online]: [RPubs - Triangular Distribution](https://rpubs.com/pac25/la1)

Zach (3 December 2021): How to Use the Triangular Distribution in R; [online]: [How to Use the Triangular Distribution in R (With Examples) - Statology](https://www.statology.org/triangular-distribution-in-r/)