Module 2 Project — Benefit-Cost Analysis of Dam Construction Projects

Sourabh D. Khot (ID 002754952)

College of Professional Studies, Northeastern University

ALY 6050: Introduction to Enterprise Analytics CRN 81180

Professor Azadeh Mobasher

June 6, 2022

Table of Contents

Problem Introduction

JET Corporation is a construction company based in the southeastern United States. In their annual planning, its management is considering two dam construction projects: The dam 1 project in southwest Georgia and the dam 2 project in North Carolina. However, it has resources for building only one project in the coming five years.

As a data analyst, I have to evaluate the costs and benefits associated with both the projects and then recommend one of them to the management with reasoning. Using subject knowledge and empirical data of the costs and benefits, I will primarily employ Monte Carlo simulation along with the concepts of statistics and probability to compute the benefit-cost ratio of the projects and arrive at a conclusion.

Problem Analysis

Through regional subject matter experts, six areas of benefits and two categories of cost of building the dam are identified and quantified (in million dollars) using historical data as given in Table 1. Since only the minimum, mode and maximum are known, the probability distribution of these variables is modeled using the 'lack of knowledge' or triangular distribution.

($ million)	*Dam 1*			*Dam 2*		
	Min	Mode	Max	Min	Mode	Max
Benefit						
Improved navigation B1	1.1	2	2.8	2.1	3	4.8
Hydroelectric power B2	8	12	14.9	8.7	12.2	13.6
Fish and wildlife B3	1.4	1.4	2.2	2.3	3	3
Recreation B4	6.5	9.8	14.6	5.9	8.7	15
Flood control B5	1.7	2.4	3.6	0	3.4	3.4
Commercial development B6	0	1.6	2.4	0	1.2	1.8
Cost						
Annualized capital cost C1	13.2	14.2	19.1	12.8	15.8	20.1
Operations & Maintenance C2	3.5	4.9	7.4	3.8	5.7	8

Table 1. Benefits and costs for the Dam #1 and #2 construction projects

For recommending a project, I will use the benefit-cost ratio as the yardstick, which is the ratio of total benefits to total costs of a particular project. Principally, for undertaking any project, its benefit-cost ratio should be greater than 1 for breaking even. Secondly, the higher the benefit-cost ratio of a project, the more sense it makes to undertake the project first because of higher benefits relative to the costs, especially when there are limited resources.

Part 1: Benefit-Cost Simulation and Observed Sample Analysis

(i) Simulation of benefit-cost ratios for Dam 1 and Dam 2

Each of the six benefits and two costs was independently simulated 10,000 times using specified triangular distribution for dam 1 and dam 2. Thus, a total of 12 sets of 10,000 independent random numbers were used. For one simulation, the total benefit was obtained as the sum of the six benefits, and the total cost was obtained as the sum of the two costs. The ratio of total benefits to total costs was computed, denoted by α_1 for dam 1, and α_2 for dam 2, each having 10,000 simulated or observed values.

(ii) Frequency Distribution of Benefit-Cost Ratio of Dam 1 (α_1) and Dam 2 (α_2)

The 10,000 observations of α_1 and α_2 are binned into 100 classes each, and their graphical frequency distribution is plotted in Figures 2 and 3 respectively. The classes are labeled using the class midpoint. These distributions are ratios of two sums of triangular distributions and, theoretically, cannot be assigned a known probability distribution. However, we can hypothesize a theoretical distribution from the visual inspection of the simulated values and test its validity, as done for α_1 in Part 2.

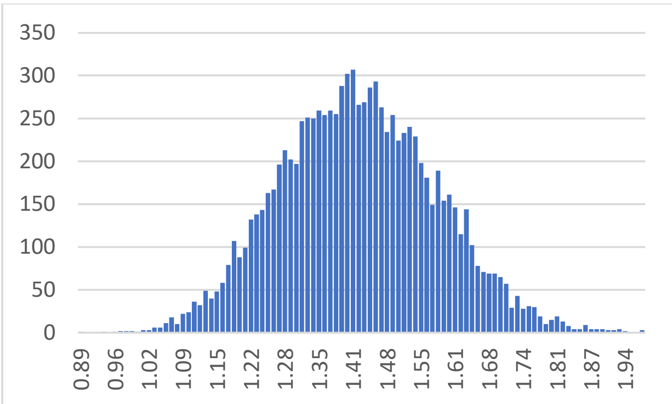

Figure 2. Observed Frequencies of α_1

In Figure 2, the distribution of the benefit-cost ratio of the dam 1 project has the most values towards the center at around 1.4, with bin frequencies tapering off towards the two ends. It visually resembles a normal distribution with few big spikes and slight skewness.

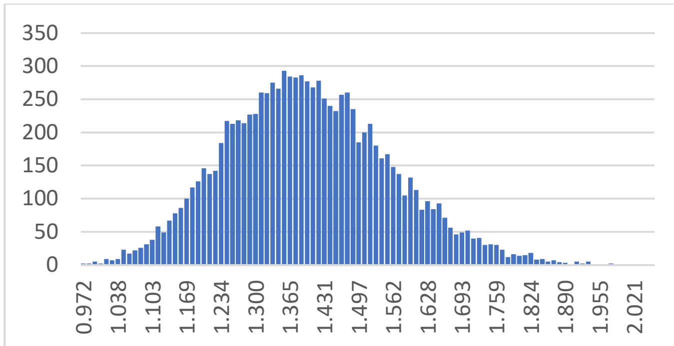

Figure 3. Observed Frequencies of α_2

Similarly, the ratio distribution of dam 2, as given in figure 3, has a visual center around 1.39 with frequencies tapering off towards the two ends. However, there is a noticeable right-skewness and short but more spikes throughout; hence, it may not be normal distribution but might potentially be a log-normal distribution that is also right-skewed.

(iii) Observed vs. Theoretical Descriptive Statistics of Benefits, Costs & Ratio

The observed mean and standard deviation (SD) for benefits, costs, and their ratio are computed from the observed values. For theoretical values, I have used the statistics formulae for triangular distribution and sum of random variables given below.

$$mean(X + Y) = mean(X) + mean(Y)$$

$$var(X + Y) = var(X) + var(Y) + 2 * cov(X, Y)$$

In other words, the mean of the sum of random variables is the sum of their means. And for the variance of sum, in our case, all random variables are independent; hence, the covariance is zero, and the variance of the sum is simply the sum of the variances.

	Dam 1	
	Observed	**Theoretical**
Mean of the Total Benefits	29.50	29.47
SD of the Total Benefits	2.33	2.31
Mean of the Total Cost	20.75	20.77
SD of the Total Cost	1.53	1.52
Mean of the Benefit-Cost Ratio (α_1)	1.43	-
SD of the Benefit-Cost Ratio (α_1)	0.15	-

Table 4. Descriptive Statistics for of Dam 1 Project Variables

For Dam 1, the observed and theoretical values are very close. Also, the observed mean benefit-cost ratio is positive, indicating that the total benefits are higher than the total cost.

	Dam 2	
	Observed	**Theoretical**
Mean of the Total Benefits	30.71	30.70
SD of the Total Benefits	2.39	2.41
Mean of the Total Cost	22.08	22.07
SD of the Total Cost	1.73	1.73
Mean of the Benefit-Cost Ratio (α_2)	1.40	-
SD of the Benefit-Cost Ratio (α_2)	0.16	-

Table 5. Descriptive Statistics for Dam 2 Project Variables

The Dam 2 observed values are also close to the theoretical values with a positive benefit-cost ratio. Compared to dam 1, dam 2 mean total benefit is more, but the mean total cost is also more. However, the dam 2 mean benefit-cost ratio (1.40) is slighter lower than Dam 1 (1.43).

Part 2: Probability Distribution Fitting for Dam 1

Fit a theoretical probability distribution on α_1 and perform a Goodness of fit test

The distribution of α_1 looks like a normal distribution but there are some spikes and very slight right skewness. Hence, I will test it for both normal and triangular distribution.

I selected triangular distribution because the observed distribution is the ratio of the sum of triangular distributions. It has a defined range of minimum and maximum, estimated mode, and slight skewness with no other knowledge. Accordingly, the parameters a, b and c are estimated, and the theoretical distribution is plotted on top of observed frequencies in Figure 6.

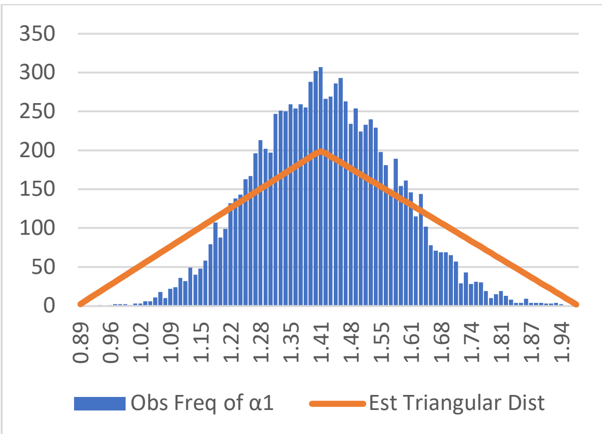

Figure 6. Fitting a triangular dist. on α_1

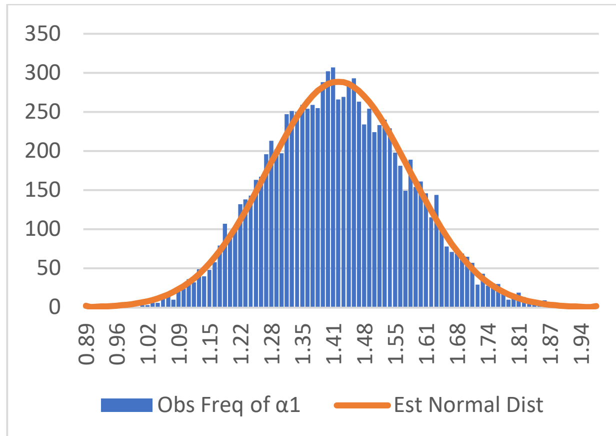

Figure 7. Fitting a normal dist. on α_1

Visually, triangular distribution fits poorly against the observed frequencies. Similarly, I fitted a normal distribution in Figure 7 because visually, the mean is almost at the center of the distribution (i.e. mean = mode = median) with frequencies tapering off towards both the ends like a natural phenomenon. It seems to fit fine, however, I have performed the chi-square goodness of fit test to confirm statistically with the following hypothesis and results in Table 8.

H_0: The proposed theoretical (triangular/normal) is a good fit on α_1

H_1: The proposed theoretical (triangular/normal) is not a good fit on α_1

	Triangular Distribution	Normal Distribution
Level of Significance	0.05	0.05
d.f. *(bins - para est - 1)*	96	97
Chi-Squared Test Statistic	4254.1	180.4
Chi-Squared p-value	0.000	0.000
Interpretation	Sufficient evidence to reject H_0	Sufficient evidence to reject H_0

Table 8. Chi-Square Goodness-of-Fit Test on α_1

Degrees of freedom for test on triangular distribution test is 96 since three parameters (a, b, c) are estimated, while that for normal is 97 since only two (mean, SD) are estimated. Table 8 compiles the goodness-of-fit test results against both triangular and normal distribution at 0.05 level of significance, where the chi-squared test statistics are 4254.1 and 180.4, respectively.

Visually normal distribution is a better fit; however, the test p-value against both the distribution is zero. Hence there is sufficient evidence to claim that the observed distribution is neither a triangular nor a normal distribution. It may follow some other distribution.

Part 3: Comparison of Benefit-Cost Ratios and Recommendations

(i) Comparison Summary of Benefit-Cost Ratios

The descriptive statistics of 10,000 simulations of benefit-cost ratios are given in Table 9. The cumulative probability of the ratio being greater than a specified number is also computed.

	Dam 1, α_1	Dam 2, α_2
Minimum	0.90	0.94
Maximum	1.99	2.07
Mean	1.43	1.40
Median	1.42	1.39
Variance	0.02	0.02
Standard Deviation	0.15	0.16
Skewness	0.16	0.32
$P(\alpha_i > 2)$	0.000	0.000
$P(\alpha_i > 1.8)$	0.009	0.008
$P(\alpha_i > 1.5)$	0.308	0.243
$P(\alpha_i > 1.2)$	0.932	0.902
$P(\alpha_i > 1)$	0.999	0.998

Table 9. Comparison Summary of Benefit-Cost Ratios

Although the minimum and maximum ratio values of dam 2 are greater than dam 1, the mean and median of dam 1 are almost always greater than that of dam 2. Also, the standard deviation or risk is less in Dam 1. Observing the cumulative probability, the probability that the benefit-cost ratio is greater than 1 [$P(\alpha_i > 1)$] is 0.999 for dam 1 and 0.998 for dam 2, which means that both the projects will break even with more than 99% confidence. Comparing other cumulative probabilities, dam 1 project is almost always better than dam 2, e.g. $P(\alpha i > 1.2)$ is 93.2% for dam 1 and 90.2% for dam 2 and so on.

(ii) Probability that Dam 1 project has ratio higher that that of Dam 2

The observed probability that $\alpha_1 > \alpha_2$ is computed by comparing the 10,000 simulations of α_1 and α_2 and dividing the count by 10,000 and the result is 55.3%. That is, 553 of 1000 times, dam 1 project will have a higher benefit-cost ratio than the dam 2. That is not very high, but still greater than 50%. Hence, with limited resources, I recommend going ahead with dam 1. My rationale for this recommendation is summarized in the next section.

Conclusion

The six benefits and two costs for the two projects were simulated 10,000 times using triangular distributions. The resultant benefit-cost ratio distribution was critically evaluated and compared to determine the better of the two projects.

The shape of the distribution of dam 1 ratio is not known to conform to any theoretical distribution as the goodness-of-test distribution rejected the hypothesis that it is a triangular or normal distribution.

Both the dam 1 and dam 2 projects had ratios of greater than 1 more than 99% of the time, with little difference in other statistics. Hence, both the projects are good, and we will get benefits from executing both of them. However, with limited resources available for only one project, I will recommend going with the dam 1 project.

Of dam 1, the expected and the median benefit-risk ratio is higher and the risk is lower than dam 2. Also, the probability that the benefit-cost ratio is higher than a particular value is always better for dam 1, for instance, the probability that the ratio being higher than 1.2 is 93.2% for dam 1 and 90.2% for dam 2. Lastly, the times that dam 1 project will outperform dam 2 project is 553 of 1000.

Hence my final recommendation is to go ahead with the dam 1 project.

.

References

APA Style Table: APA.org. (n.d.). Retrieved from https://apastyle.apa.org/style-grammar-guidelines/tables-figures/tables

Canvas Module 2 Project: Analysis of a Betting Strategy in Sports. (n.d.). Retrieved from https://northeastern.instructure.com/courses/110059/assignments/1346042

Evans, J. R. (2013). *Statistics, data analysis, and decision modeling.* New York: Pearson.

https://www.probabilitycourse.com/chapter6/6_1_2_sums_random_variables.php. (n.d.). *probabilitycourse.com: 6.1.2 Sums of Random Variables*. Retrieved from https://www.probabilitycourse.com/chapter6/6_1_2_sums_random_variables.php

Microsoft Support: Format text or numbers as superscript or subscript. (n.d.). Retrieved from https://support.microsoft.com/en-us/office/format-text-or-numbers-as-superscript-or-subscript-3649411b-adf4-483e-b0e8-7b844605da74

trumpexcel.com: How to Use Multiple Criteria in Excel COUNTIF and COUNTIFS Function. (n.d.). Retrieved from https://trumpexcel.com/multiple-criteria-in-excel-countif/

<End of Report>