A Benefit Cost Analysis of Two Dams

Molly Szeman

The Roux Institute, Northeastern University

ALY6050: Introduction to Enterprise Analytics

Prof. Andrew Kinnley

November 16, 2023

Part 1: Introduction, Benefit Cost Ratios, and Frequency Distributions

This analysis aims to recommend the building of one of two proposed dams that will provide the most economic benefit to the surrounding community. Experts have compiled the minimum, maximum, and mode of possible economic costs and benefits for Dam #1, located in Georgia, and Dam #2, located in North Carolina, as seen in Table 1. Six categories of benefits (improved navigation, hydroelectric power, fish and wildlife, recreation, flood control, and commercial development) and two categories of costs (annualized capital cost and operations and maintenance) were considered, with the values of benefits and costs given in millions of dollars ($).

Figure 1: Benefits and Costs of Two Dams

Dam #1				Dam #2			
Benefits:	Minimum	Mode	Maximum	Benefits:	Minimum	Mode	Maximum
Improved navigation	1.1	2	2.8	Improved navigation	2.1	3	4.8
Hydroelectric power	8	12	14.9	Hydroelectric power	8.7	12.2	13.6
Fish and wildlife	1.4	1.4	2.2	Fish and wildlife	2.3	3	3
Recreation	6.5	9.8	14.6	Recreation	5.9	8.7	15
Flood control	1.7	2.4	3.6	Flood control	0	3.4	3.4
Commercial development	0	1.6	2.4	Commercial development	0	1.2	1.8
Costs:	Minimum	Mode	Maximum	Costs:	Minimum	Mode	Maximum
Annualized capital cost	13.2	14.2	19.1	Annualized capital cost	12.8	15.8	20.1
Operations & Maintenance	3.5	4.9	7.4	Operations & Maintenance	3.8	5.7	8

A triangular distribution was used to estimate the average cost, average benefit, and the average cost benefit ratio for each dam. These means were calculated using the formula:

$E(x) = (a+b+c)/3$, where a=min, b=max, and c=mode.

The sums of each category were added to determine a total theoretical benefit and cost (in millions of dollars) for each dam. The benefit cost ratio for each dam was calculated by dividing the total theoretical benefit by the total theoretical cost.

The highest potential benefit for each dam was hydroelectric power (11.633 for Dam #1 and 11.5 for Dam #2), and the highest cost for each dam was annualized capital cost (15.5 and 16.233) (Figure 2). Most categories were within 1 of each other (representing $1 million), but Dam #2 had significantly higher economic benefit from improved navigation and and fish and wildlife, while Dam #1 held a small edge in the other four categories (Figure 2). The costs were estimated to be higher in both categories for Dam #2 (Figure 2). Figure 2 shows that Dam #2 has a higher theoretical benefit (30.7) and cost (22.067) than Dam #1 (29.467 and 20.767), but a lower benefit cost ratio (1.391 for Dam #2 to 1.42 for Dam #1).

Figure 2: Theoretical Means of Costs and Benefits of Two Dams

Benefits:	**Dam #1**	**Dam #2**
Improved navigation	1.967	3.300
Hydroelectric power	11.633	11.500
Fish and wildlife	1.667	2.767
Recreation	10.300	9.867
Flood control	2.567	2.267
Commercial development	1.333	1.000
Total Benefit:	29.467	30.700
Costs:		
Annualized capital cost	15.500	16.233
Operations & Maintenance	5.267	5.833
Total Cost:	20.767	22.067
Benefit Cost Ratio:	1.420	1.391

To further investigate the potential impacts of these projects and in order to select the project most likely to have the highest economic impact, 10,000 simulated outcomes of possible benefits and costs for each of the eight categories were created for each dam.

This was done by using the inversion method of creating random outcomes from a probability distribution. The formula

if $r \le K$, $x = a + \sqrt{rM}$ or if $r > K$, $x = b - \sqrt{(1-r)N}$

was used

given $K = (c-a) / (b-a)$, $M = (b-a)(c-a)$, and $N = (b-a)(b-c)$

Where K calculates the proportion between the mode to the minimum and the maximum to the minimum. M is related to the area of the left part of the triangle, and N is related to the area of the right part of the triangle. The randomly generated number r is compared to K and determines if the area of the left or right side of the triangle is used.

Additionally, a cost benefit ratio was calculated for each of the 10,000 trials by dividing the theoretical average benefit by the theoretical average cost.

In the simulated outcomes, Dam #1 had a minimum cost benefit ratio of 0.924, while Dam #2's minimum value was 0.948, which is 0.024 higher (Figure 3). Dam #2 also had a higher maximum (2.073) than Dam #1 (2.001), and a larger range between minimum and maximum values (Figure 3).

Figure 3: Descriptive Statistics of Cost Benefit Ratio of Two Dams

	Dam #1	Dam #2
Minimum	0.924	0.948
Maximum	2.001	2.073
Range	1.077	1.126

These statistics highlight only two extreme outcomes for each dam out of the 10,000 trials, so further analysis into the likely outcomes must be done to determine which dam is more likely to have a greater economic benefit for its region.

The simulated outcomes for each dam were grouped into 100 bins each. 100 bins were used because 100 is the square root of 10,000, and that was used as a starting point. When the frequencies were calculated for each bin, in Dam #1, there were 3 bins containing 0 outcomes and 20 bins with fewer than 5 outcomes. For Dam #2, there were 9 bins with 0 outcomes and 21 bins with fewer than 5 outcomes. While reducing the number of bins was considered, since there were relatively few 0s and most bins have 5 or more outcomes, 100 bins were used to create the frequency distributions. Figure 4 and 5 show the frequency distributions of Dam #1 and Dam #2.

Figure 4:

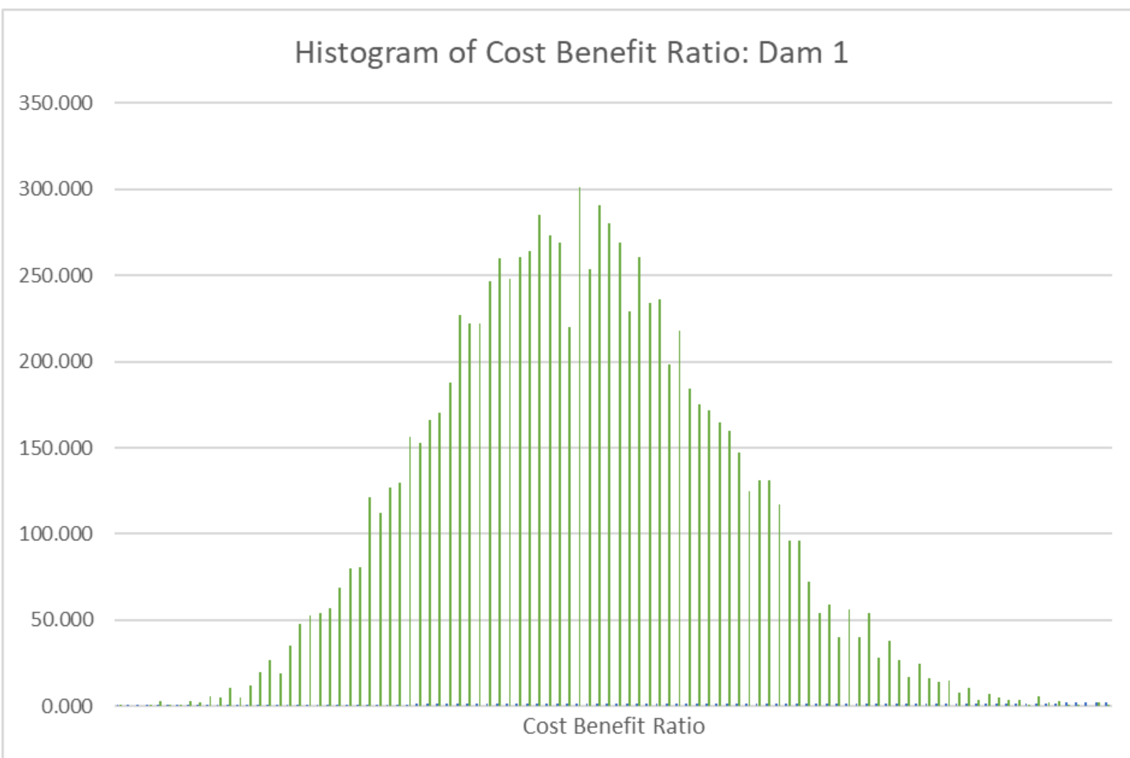

Figure 5:

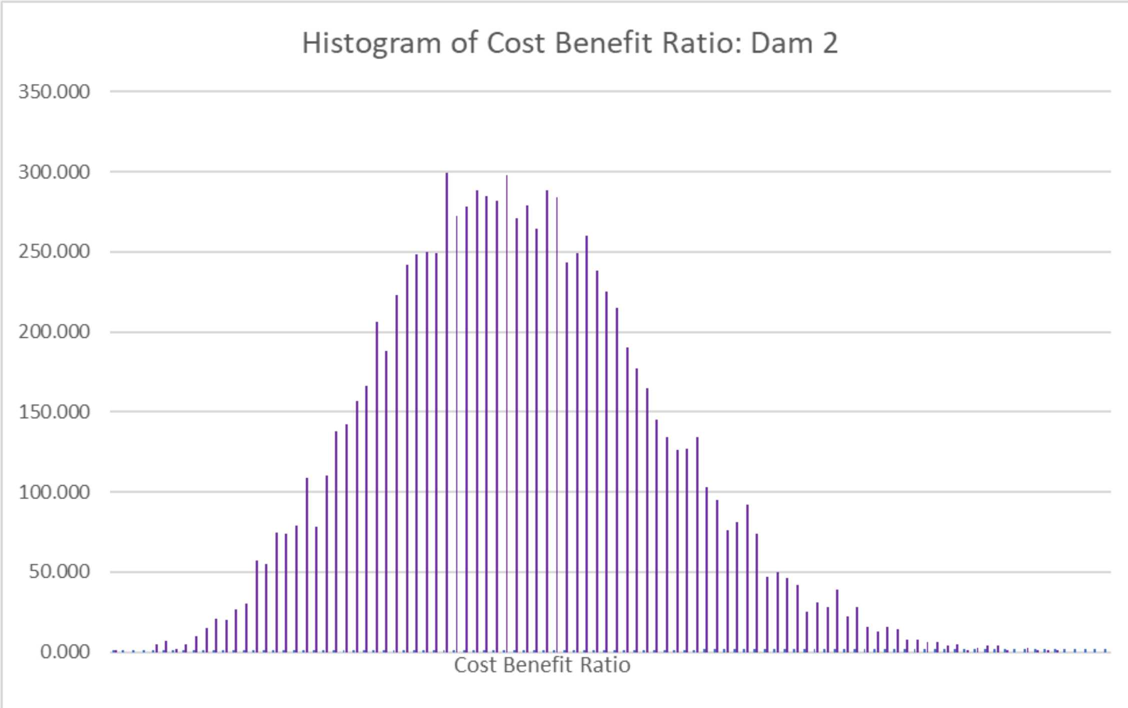

Figure 6 shows the means and standard deviations for the theoretical distributions and the simulated trials. For both Dams 1 and 2, the observed statistics closely resemble the theoretical values (Figure 6). In Dam #1, the theoretical mean of total benefits was 0.034 larger than the observed value, and the theoretical mean of the total cost was 0.017 larger than the observed value. In Dam #2, the observed mean of total benefits was 0.018 larger than the theoretical, and the theoretical mean of total cost was 0.011 larger than the observed. Both dams had similarly close standard deviations of cost and benefit.

Figure 6: Means and Standard Deviations of Theoretical and Observed For Dams 1 and 2

Dam 1	Observed	Theoretical
Mean of the Total Benefits	29.433	29.467
SD of the Total Benefits	2.326	2.307
Mean of the Total Cost	20.750	20.767
SD of the Total Cost	1.532	1.521
Mean of the Benefit-cost Ratio	1.426	X
SD of the Benefit-cost Ratio	0.154	X

Dam 2	Observed	Theoretical
Mean of the Total Benefits	30.718	30.700
SD of the Total Benefits	2.415	2.410
Mean of the Total Cost	22.056	22.067
SD of the Total Cost	1.731	1.727
Mean of the Benefit-cost Ratio	1.401	X
SD of the Benefit-cost Ratio	0.155	X

Part 2: Goodness of Fit Test

Next, a Chi-squared Goodness of Fit test was used to determine if the simulated distribution matches a theoretical distribution.

- H_0: There are no significant differences between the simulated data and the theoretical distribution.
- $H_{1:}$ There are significant differences between the simulated data and the theoretical distribution.

A normal distribution and a triangular distribution were considered based on the visual appearance of the histograms and the fact that the costs and benefits were given in a triangular distribution. A normal distribution was selected as a first attempt due to the maximum number of occurrences in Dam #1 (301) being located in the range of 1.42-1.43, which is close to the center of the distribution (1.463) in the 47th bin.

A theoretical probability for the normal distribution was calculated using the mean (1.426) and standard deviation (0.154) of the observed trials. This was created using the =NORM.DIST function in Excel. The expected frequency of the normal distribution was calculated by multiplying the expected frequency by 10,000. Next, the Chi-squared test statistic was calculated in each bin using the formula:

$= (\text{Expected - Observed})^2 / \text{Expected}$

The total Chi-squared test statistic was summed, equalling 173.847. Using that value and degrees of freedom = 97 (calculated as n - 1 - p, where n is the number of bins, 100, and p is the number of estimated parameters, 2), the Chi-squared p-value was calculated using the =CHI.SQ.DIST.RT function in Excel. The chi-squared p-value = 0.000005.

Since the p-value is less than the alpha (0.05), we reject the null hypothesis that there is no difference between the simulated data and the expected normal distribution. Further analysis should be done to determine if the bins should be reduced or if the distribution fits another distribution, such as triangular, or another distribution not initially considered here.

Part 3: Conclusions and Recommendations

After determining the theoretical benefit cost ratio and analyzing 10,000 simulated trials for each dam, this report recommends proceeding with Dam #1 as the project that is more likely to provide financial benefit to the region.

Figure 7: Comparative Statistics of Dam #1 and Dam #2

	$\boldsymbol{\alpha}_1$	$\boldsymbol{\alpha}_2$
Minimum	0.924	0.948
Maximum	2.001	2.073
Mean	1.426	1.401
Median	1.423	1.393
Variance	0.024	0.024
Standard Deviation	0.154	0.155
SKEWNESS	0.187	0.327
$P(\boldsymbol{\alpha}_i > 2)$	0.0001	0.0002
$P(\boldsymbol{\alpha}_i > 1.8)$	0.0102	0.0099
$P(\boldsymbol{\alpha}_i > 1.5)$	0.308	0.250
$P(\boldsymbol{\alpha}_i > 1.2)$	0.9335	0.9088
$P(\boldsymbol{\alpha}_i > 1)$	0.9993	0.9992
$P(\boldsymbol{\alpha}_1 > \boldsymbol{\alpha}_2)$	0.547	

While Dam #2 had higher minimum and maximum simulated values, indicating it might produce more economic output, the mean and median of the trials were both higher in Dam #1, suggesting the outcomes more likely to occur favored Dam #1 as more beneficial. Dam #2 also had greater skewness in its frequency distribution, although both dams skewness were positive and relatively moderate (Evans, 2013; Figure 7). Greater skewness suggests the results may contain more positive outliers and may have produced misleading results (Evans, 2013; Figure 6).

Dam #2 had a slightly higher likelihood of producing a benefit cost ratio of greater than 2, but Dam #1 had a higher probability of benefit cost ratios of greater than 1.8, 1.5, 1.2, and 1 (Figure 6). Additionally, when each of the 10,000 trials from Dam #1 were compared to its counterpart in Dam #2, Dam #1 had a higher benefit cost ratio 54.7% of the time (Figure 7). These statistics indicate that Dam #1 is the better choice because it is more reliably a better performer than Dam #2, despite Dam #2 having higher maximum potential benefit (Figure 7).

References

Chatgpt. ChatGPT. (2023, November 16). https://openai.com/chatgpt

Evans, J. (2013). *Statistics, data analysis, and decision modeling* (5th ed.). Pearson.