ASSIGNMNET 2

SAHL GAWANDE

ALY6050: Introduction to Enterprise Analytics

Professor: Zhi Richard He

June 8, 2022

Abstract

As a response to your inquiry, I've embarked on a cost-benefit assessment for the firm, utilizing a triangular distribution model. This involved generating random numbers to forecast the most frequent outcomes from a specific sequence. The context of this assessment is based on a corporation related to the construction of two dams in states like North Carolina and Georgia. In the process, I have evaluated statistical data such as observed and theoretical distributions. On visualizing the frequency distribution, the gamma distribution became apparent. In addition, I performed a chi-square test to evaluate the goodness of fit and determined the P-value to interpret the findings and suggest which construction should be taken into account.

Introduction

For the purpose of assessing the dual-dam construction initiative in Georgia and North Carolina, I carried out a cost-benefit analysis for JET Corporation. The firm has recognized six potential benefits from this project, which include improved flood control, enhanced navigation, fisheries, hydroelectric power, wildlife conservation, recreational opportunities, flood mitigation, and local business expansion. Additionally, the analysis takes into account the cumulative capital cost over a span of 30 years along with operational and maintenance expenses. A benefit-cost ratio is formulated by dividing the total benefits by the actual costs. When the ratio exceeds one, it signifies that the benefits surpass the costs. The higher this ratio, the more likely the project is to be selected over competing ventures. Analysis

A simulation with 10,000 benefit-cost ratios for dams 1 and 2 should be run. These simulations are separate from one another. The following two tables, for dam1 and dam2, were provided in the question.

Dam #1: Benefits & Costs

Benefit	Estimate		
	Minimum	*Mode*	*Maximum*
Improved navigation **B1**	1	4	5
Hydroelectric power **B2**	6	11	15
Fish and wildlife **B3**	0	3	6
Recreation **B4**	1	5	10
Flood control **B5**	0	3	7
Commercial development **B6**	2	5	9

Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	10	12	17
Miscellaneous costs **C2**	0	1	2
Operations & Maintenance **C3**	1	3	5

Dam # 2: Benefits & Costs

Benefit	Estimate		
	Minimum	*Mode*	*Maximum*
Improved navigation **B1**	5	7	8
Hydroelectric power **B2**	5	8	9
Fish and wildlife **B3**	1	2	3
Recreation **B4**	2	4	7
Flood control **B5**	2	5	6
Commercial development **B6**	0	2	5

Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	7	8	9
Miscellaneous costs **C2**	1	2	3
Operations & Maintenance **C3**	3	4	5

Fig 1 : Benefit & Cost for Dam1 Fig 2 : Benefit & Cost for Dam2

I initially incorporated the triangular probability distribution in my assessments. Given the scarce data available, this distribution is often referred to as the distribution of ignorance. In the triangular distribution, the minimum value, maximum value, and mode are represented as a, b, and c respectively, as depicted in the diagram below. Notably, the area beneath the triangle equals 1.

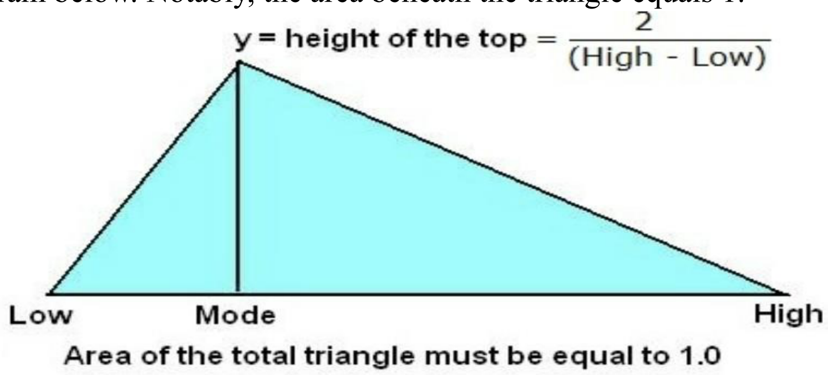


Fig 3: Triangle Distribution

The formula presented below is used to compute the probability distribution function for a point that lies within the range of 'a' to 'b'. b.

$$f(x) = \begin{cases} \frac{2}{(b-a)(c-a)}(x-a) & \text{if} \quad x \le c \\ \frac{2}{(b-a)(b-c)}(b-x) & \text{if} \quad x > c \end{cases}$$

Fig 4: Triangle Distribution Formula

To determine the Probability Distribution Function (PDF), Cumulative Probability Function (CPF), Random Number Generation (RNG), mean, and variance, I've methodically computed these values in a step-by-step process.

Benefits:	Triangular Distribution:			Triangular Distribution:			Triangular Distribution:		
Dam I:	left end - a	peak - c	right end - b	c-a	b-a	b-c	K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
Improved navigation B1	1	4	5	3	4	1	0.75	12	4
Hydroelectric power B2	6	11	15	5	9	4	0.56	45	36
Fish and wildlife B3	0	3	6	3	6	3	0.50	18	18
Recreation B4	1	5	10	4	9	5	0.44	36	45
Flood control B5	0	3	7	3	7	4	0.43	21	28
Commercial development B6	2	5	9	3	7	4	0.43	21	28
Costs:									
Dam I:	left end - a	peak - c	right end - b	c-a	b-a	b-c	K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
Annualized capital cost C1	10	12	17	2	7	5	0.29	14	35
Miscellaneous costs C2	0	1	2	1	2	1	0.50	2	2
Operations & Maintenance C3	1	3	5	2	4	2	0.50	8	8

Fig 5: precalculated values for Dam1

Benefits:

Dam 2:	Triangular Distribution: left end - a	peak - c	right end - b	Triangular Distribution: c-a	b-a	b-c	Triangular Distribution: K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
Improved navigation B1	5	7	8	2	3	1	0.67	6	3
Hydroelectric power B2	5	8	9	3	4	1	0.75	12	4
Fish and wildlife B3	1	2	3	1	2	1	0.50	2	2
Recreation B4	2	4	7	2	5	3	0.40	10	15
Flood control B5	2	5	6	3	4	1	0.75	12	4
Commercial development B6	0	2	5	2	5	3	0.40	10	15

Costs:

Dam 2:	left end - a	peak - c	right end - b	c-a	b-a	b-c	K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
Annualized capital cost C1	7	8	9	1	2	1	0.50	2	2
Miscellaneous costs C2	1	2	3	1	2	1	0.50	2	2
Operations & Maintenance C3	3	4	5	1	2	1	0.50	2	2

Fig 6: precalculated Values for Dam 2

RNG (Random Number generation)

if $r \leq K$, then $x = a + \mathrm{sqrt}(r*A)$

if $r > K$, then $x = b - \mathrm{sqrt}((1-r)*B)$

Fig 7: Equation for Random number generation

Mean & Variance

Mean: $\mu = \frac{a+b+c}{3}$

Variance: $\sigma^2 = \frac{a^2+b^2+c^2-ab-ac-bc}{18}$

Fig 8: Equations for mean and variance

Leveraging a triangular distribution, we're now able to calculate the theoretical mean and variance values for both Dam1 and Dam2, considering both benefits and costs.

Theoretical Mean: E(X) = = (a+b+c)/3	Theoratical variance
3.333	0.722222222
10.667	3.388888889
3.000	1.5
5.333	3.388888889
3.333	2.055555556
5.333	2.055555556

Theoretical E(X)	
13.000	2.166666667
1.000	0.166666667
3.000	0.666666667

Theoretical Mean: E(X) = = (a+b+c)/3	
6.667	0.388888889
7.333	0.722222222
2.000	0.166666667
4.333	1.055555556
4.333	0.722222222
2.333	1.055555556

Theoretical E(X)	
8.000	0.166666667
2.000	0.166666667
4.000	0.166666667

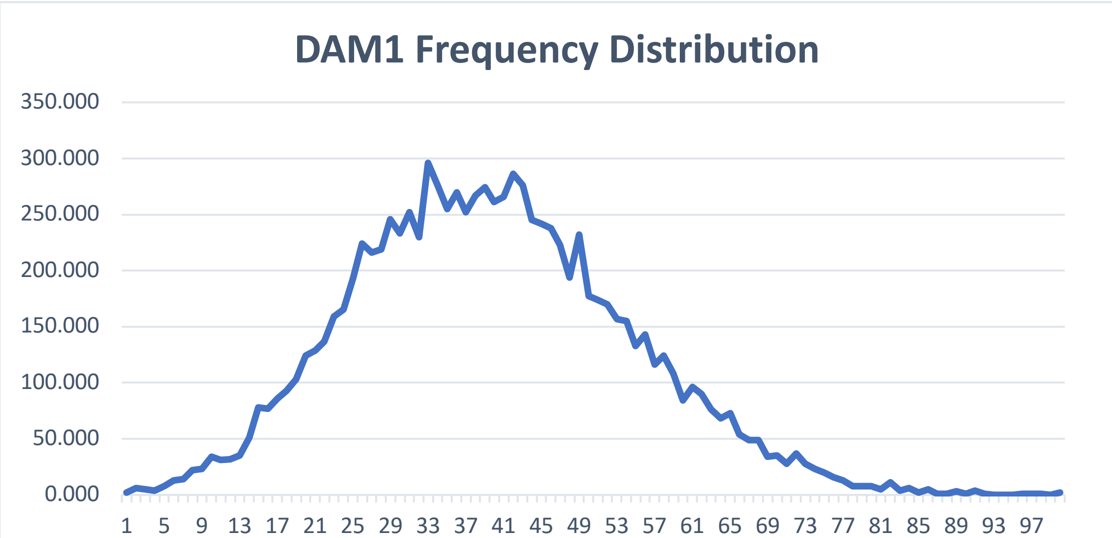
Fig 8: Theoretical Values for Dam 1 and Dam 2

By generating random numbers, we can now empirically calculate all benefits for 10,000 instances. After generating a random number for each benefit, I've summed up every benefit for a total of 10,000 values. The two distinct cost values are processed in a similar manner. After computing each of these values, I was able to ascertain the benefit-cost ratio for both dams. Subsequently, within the headers of alpha1 and alpha2, I calculated the minimum, maximum, and range of the benefits for both dams. These values are then used to calculate frequency distribution parameters, such as class left, class right, class midpoint, and class frequency.

	ALPHA1	ALPHA2
MIN	0.735	1.365
MAX	2.964	2.619
RANGE	2.229	1.255
Class/Bins	100.000	100.000
Class width	0.022	0.013
COUNT	10000.000	10000.000

Fig 9: Statistical values both dams

After plotting the class frequency, I have a curve that depicts the gamma distribution, as demonstrated in Figure 10. Given that the values extend beyond the range of zero to one, which is the range for the beta distribution, the gamma distribution is represented in this Figure 10.



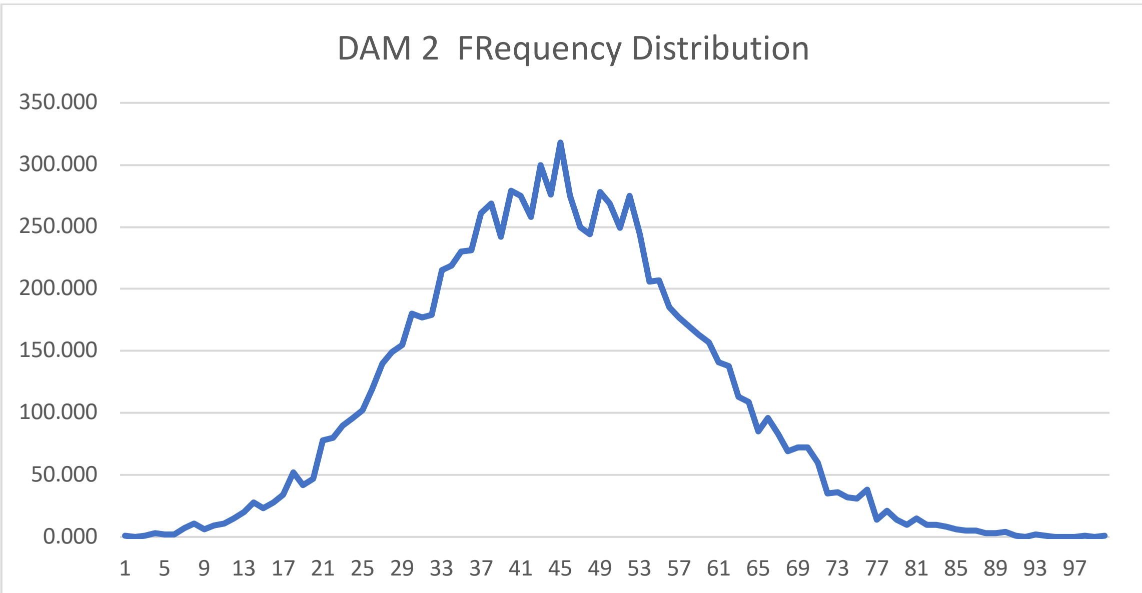


Fig 10: Frequency distribution of Dam 1 and Dam 2

The outcomes derived from computing the observed and theoretical values are displayed in Figure 11. For both Dam1 and Dam2, the observed values align closely with the theoretical predictions.

Dam 1	Observed	Theoretical
Mean of the Total Benefits	29.253	31.000
SD of the Total Benefits	3.889	3.621
Mean of the Total Cost	16.982	17.000
SD of the Total Cost	1.733	1.732
Mean of the Benefit-cost Ratio	1.741	X
SD of the Benefit-cost Ratio	0.296	X

Dam 2	Observed	Theoretical
Mean of the Total Benefits	26.997	27.000
SD of the Total Benefits	2.033	2.028
Mean of the Total Cost	14.017	14.000
SD of the Total Cost	0.711	0.707
Mean of the Benefit-cost Ratio	1.931	X
SD of the Benefit-cost Ratio	0.177	X

Figure 11: Calculation of observed and theoretical values of Dam1 and Dam2

applying the variables to identify the alpha 1 or dam1 distribution with the best match. The chi-square test is a way to determine whether the null hypothesis should be accepted or rejected. The alternative hypothesis is that the gamma distribution is not the best fit for our hypothesis, contrary to what the null hypothesis claims. The P-value in our situation is 0.0, which means
$p < 0.05$. Thus we can reject the null hypothesis is thus a result of sufficient data. As I have considered random values, the chi-squared P-value has fluctuated, but on average it will produce the same result.

Chi-squared Test Statistic:	**#DIV/0!**
Chi-squared P-value:	**0.000**

Figure 12: chi-square test values

Moreover, I completed the table with figures corresponding to alpha 1 and alpha 2 (or Dam1 and Dam2), which included the maximum, minimum, mean, median, variance, among other metrics of the benefit-cost ratio. The following chart indicates that P-values for Dam1 consistently outperform alpha values for Dam2. Given its superior benefit-cost ratio, Dam1 in Southwest Georgia should take precedence over Dam2 in North Carolina in terms of construction priorities.

	α_1	α_2
Minimum	0.938	1.367
Maximum	2.984	2.636
Mean	1.741	1.931
Skewness	0.308339227	0.136107928
Kurtosis	0.005337503	-0.07638774
Standard Deviation	0.296188426	0.177345418
$P(\alpha_i > 2.25)$	0.0527	0.0401
$P(\alpha_i > 2.00)$	0.190	0.344
$P(\alpha_i > 1.75)$	0.4689	0.8437
$P(\alpha_i > 1.50)$	0.7835	0.9953
$P(\alpha_i > 1.25)$	0.9622	1
$P(\alpha_1 > \alpha_2)$	0.2883	

Fig 13: Comparing the multiple P-values

Conclusion

In this study, I initially computed the theoretical values utilizing a triangular distribution, followed by the calculation of experimental values, and subsequently compared these for both Dam1 and Dam2. Upon crafting a frequency distribution graph, I observed that the distribution curve mirrored that of a Gamma Distribution. To verify this, I employed the Chi-square test, with the null hypothesis being that the Gamma distribution is suitable in this context, and the alternative hypothesis asserting the contrary.

In scrutinizing the class frequency, all the values were found to exceed the 0 to 1 range, and the Chi-square P-value held a significance level above 0.05. Consequently, due to insufficient evidence, the null hypothesis was not dismissed. Moreover, post-calculation, the observed and theoretical values for both Dam1 and Dam2 were found to be equivalent.

Further, having obtained the simulation outcomes for alpha1 (Dam1) and alpha2 (Dam2), I concluded that Dam1 presents a more favorable benefit-cost ratio.

References

1) 1.3.6.7.4. Critical Values of the Chi-Square Distribution. (n.d.). Chi Square. https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm

2) Triangular Distribution -- from Wolfram MathWorld. (n.d.). Triangular Distribution. https://mathworld.wolfram.com/TriangularDistribution.html