MODULE TWO PROJECT

BENEFIT - COST ANALYSIS OF DAM CONSTRUCTIONS

Submitted By : **HARSHIT GAUR**
MASTER OF PROFESSIONAL STUDIES IN ANALYTICS
ALY 6050 : INTRODUCTION TO ENTERPRISE ANALYTICS
JANUARY 24, 2022

Submitted To : **PROF. RICHARD HE**

ABSTRACT

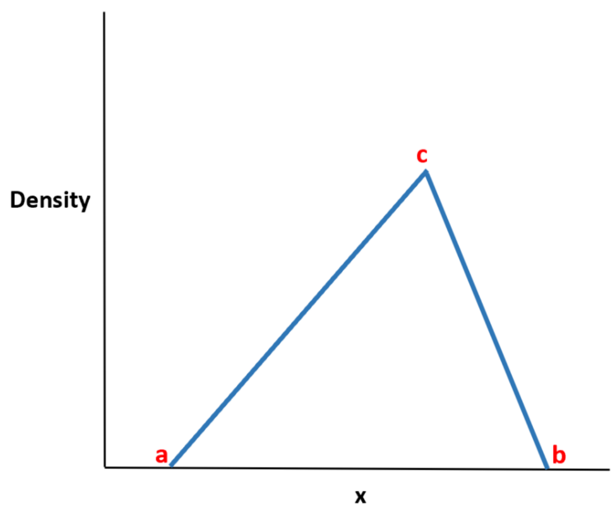
Data analytics is a discipline that focuses on obtaining meaningful and useful information from the large amount of data. It includes processes like data gathering, organisation, and storage, as well as management procedures and data analysis, strategies, and tools. The basic goal of data analytics is to use statistical analysis to uncover patterns and address problems in data.

Monte Carlo methods are the sets of algorithms which utilise inferential statistics in order to or for mimicking the behaviours of complex systems or/and probabilistic events. They replicate the physical processes which are typically time-consuming to set it up, too expensive to perform repeatedly, or both. It's a program for simulating and modelling probabilistic real-world processes.

Random number generation is a method of generating a sequence of symbols or numbers that cannot be realistically or in any other way anticipated better than by chance, usually using a random number generator (RNG). This indicates the specific outcome sequence will have some patterns that can be detected in hindsight but are unforeseeable in the future.

The **Triangular Distribution** is one of the continuous kind of probability distributions with a triangle-shaped or formed probability density function which is defined by 3 different values:

- The minimum value, *a*
- The maximum value, *b*
- The peak value, *c*



INTRODUCTION

In this project, we will be performing *Benefit-Cost Analysis* of projects available to a corporation. In this analysis, the estimations are made in different categories for both the annualised benefits & the annualised costs derived from / belonging to projects of a big corporation.

The **Benefit-Cost Ratio** of the project is produced when the total benefits are divided with the total costs involved. The benefit-cost ratio which is greater or larger than 1.0 signifies that all the benefits outweigh the costs. Therefore, the higher or larger the benefit-cost ratio of a project goes, the more chances are there for it to be selected over the projects which have lower benefit-cost ratios.

PROBLEM STATEMENT

The JET Corporation is surveying the development of two dam projects, one in southwest Georgia area (Dam #1) and the other in North Carolina area (Dam #2). Further developed route, fish and natural life, hydroelectric power, flood control, entertainment, and business improvement of the area are among the six areas of advantage referred to by the enterprise. Furthermore, for each kind of advantage, there are three appraisals accessible: a base conceivable worth, a most probable worth (i.e., a mode or top), and a greatest conceivable worth. The general capital expense, annualized north of 30 years (at a rate specified by the public authority and the lenders), and the yearly tasks and support costs have been perceived as two classifications associated with a development undertaking of this sort.

We will be performing various steps in order to achieve our goal of a proper benefit-cost analysis and answer the questions asked with regards to this problem statement.

The benefits and costs in millions of dollars for Dam #1 and Dam #2 construction projects are below.

Benefit	**Estimate**		
	Minimum	**Mode**	**Maximum**
Improved Navigation **B1**	1	2.2	2.5
Hydroelectric Power **B2**	8	11	15
Fish and Wildlife **B3**	1.4	1.4	2.2
Recreation **B4**	6.5	9.9	15.6
Flood Control **B5**	1.7	2.5	3.6
Commercial Development **B6**	0	1.6	2.4

Table 1.1: Benefits of Dam #1 construction project.

Cost	**Estimate**		
	Minimum	**Mode**	**Maximum**
Annualized Capital Cost **C1**	13.1	14.1	19.1
Operations & Maintenance **C2**	3.5	5	7.6

Table 1.2: Costs of Dam #1 construction project.

Benefit	Estimate		
	Minimum	Mode	Maximum
Improved Navigation **B1**	2.2	3	4.5
Hydroelectric Power **B2**	8.7	12.1	13.6
Fish and Wildlife **B3**	2.3	3	3
Recreation **B4**	5.9	8.6	15
Flood Control **B5**	0	3.3	3.3
Commercial Development **B6**	0	1.3	1.6

Table 1.3: Benefits of Dam #2 construction project.

Cost	Estimate		
	Minimum	Mode	Maximum
Annualized Capital Cost **C1**	12.6	15.9	20.2
Operations & Maintenance **C2**	3.6	5.7	8.1

Table 1.4: Costs of Dam #2 construction project.

Part 1

I. Perform simulations for both of the Dams (#1 and #2) of 10,000 benefit-cost ratios.

The triangular distribution is one of the continuous kind of probability distributions with lower limit - a, mode - c, upper limit - b, and in probability theory and statistics, where

- $a < b$,
- $a \leq c \leq b$

The probability distribution function (PDF) of the triangular random variables is calculated using the below function :

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

Figure 1.1: Probability Distribution Function (PDF) of triangular random variables.

We will utilize Random Number Generations of the Triangular random kind of variables to the estimates of the benefits & the costs of both of the construction projects.

Random Number Generation of Triangular Random Variables:

The CDF expression given by formula (5) can be used to generate random values according to a specific triangular distribution. In this method, first a standard uniform random value r is created. This value is then used as a cumulative probability and replaces $F(x)$ in formula (5). The formula is then solved for the random variable x. The following rule describes this random number generation:

$$x = \begin{cases} a + \sqrt{r\,(b-a)(c-a)} & \text{if } r \leq \frac{c-a}{b-a} \\ b - \sqrt{(1-r)(b-a)(b-c)} & \text{if } r > \frac{c-a}{b-a} \end{cases} \qquad (6)$$

Figure 1.2: Random Number Generation of Triangular Random Variables.

We've used the Excel worksheet to calculate the computations used in the formulae above and contained them in cells for further usage.

Triangular Distribution:			Triangular Distribution:		
c-a	b-a	b-c	K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
1.2	1.8	0.6	0.67	2.16	1.08
3	7	4	0.43	21	28
0	0.8	0.8	0.00	0	0.64
3.4	9.1	5.7	0.37	30.94	51.87
0.8	1.9	1.1	0.42	1.52	2.09
1.6	2.4	0.8	0.67	3.84	1.92

c-a	b-a	b-c	K = (c-a) / (b-a)	M =(b-a) (c-a)	N =(b-a) (b-c)
1	6	5	0.17	6	30
1.5	4.1	2.6	0.37	6.15	10.66

Figure 1.3: Calculations used in the formulae.

Theoretical Mean: E(X) = = (a+b+c)/3	Theoretical Variance: VAR(X) = (a^2+b^2+c^2-ab-bc-ca) / 18	Theoretical Standard Deviation: SD(X) = sqrt(VAR(X))
2.000	0.140	0.374
11.333	2.056	1.434
1.667	0.036	0.189
10.667	3.524	1.877
2.600	0.152	0.389
1.333	0.249	0.499

Theoretical E(X)	Theoretical Variance: VAR(X)	Theoretical Std Deviation: SD(X)
15.433	1.722	1.312
5.367	0.717	0.847

Figure 1.4: Calculations of Theoretical Mean, Variance, and Standard Deviation.

1. The formula for Theoretical Mean, E(X) = (a + b + c) / 3

2. The formula we have used for Theoretical Variance is -
   VAR(X) = (a^2 + b^2 + c^2 - a*b - b*c - c*a) / 18

3. For the Theoretical standard Deviation, the formula used is -
   SD(X) = SQRT(VAR(X))

4. We've used an excel built-in function **RAND()** to generate random variables to be used in formula for benefits and costs triangular random variables.

AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
				Total Benefits	Costs				Total Costs
r_B5	B5	r_B6	B6		r_C1	C1	r_C2	C2	
0.7497	2.88	0.7895	1.76	27.83	0.1907	14.17	0.6389	6.10	20.27
0.9482	3.27	0.5656	1.47	31.89	0.4736	15.13	0.5024	5.45	20.57
0.4282	2.51	0.0718	0.53	27.89	0.6464	15.84	0.7235	4.45	20.30
0.5000	2.58	0.8570	1.88	32.12	0.5229	15.32	0.4917	6.37	21.68
0.0118	1.83	0.4231	1.27	31.20	0.9990	18.93	0.2340	4.70	23.63
0.7167	2.83	0.3386	1.14	32.38	0.7258	16.23	0.3883	4.94	21.18
0.9630	3.32	0.7209	1.67	30.16	0.2055	14.22	0.2491	4.74	18.96
0.0811	2.05	0.4726	1.35	30.81	0.1258	13.97	0.2960	4.85	18.82
0.7586	2.89	0.9487	2.09	29.94	0.9343	17.70	0.6829	6.86	24.56
0.9928	3.48	0.3129	1.10	34.76	0.1997	14.20	0.7993	4.89	19.09
0.7926	2.94	0.2613	1.00	30.08	0.3268	14.61	0.1035	4.30	18.90
0.2678	2.34	0.2573	0.99	28.84	0.3642	14.73	0.2333	4.70	19.43
0.3304	2.41	0.6061	1.53	30.25	0.1308	13.99	0.9935	5.55	19.54

Figure 1.5: Calculations of Theoretical Mean, Variance, and Standard Deviation.

The Total Benefits and Total Costs are the summations of simulations belonging to the benefits and costs respectively.

Now, the Benefit-Cost Ratios of DAM - #1 and DAM - #2 construction projects has been calculated by dividing Total Benefits by Total Costs.

Dam 1
Benefit-cost Ratio
1.373
1.550
1.374
1.481
1.321
1.529
1.591
1.637
1.219
1.820
1.591
1.484
1.548
1.266
1.776

Figure 1.6: Benefit-Cost Ratio of Dam #1 construction project

Dam 2
Benefit-cost Ratio
1.554
1.479
1.154
1.606
1.203
1.500
1.136
1.541
1.239
1.428
1.452
1.669
1.342
1.109
1.469
1.754

Figure 1.7: Benefit-Cost Ratio of Dam #2 construction project

II. Tabular and Graphical Frequency distribution for both the ratios.

	DAM 1 (α1)	DAM 2 (α2)
Min	0.918	0.894
Max	2.044	2.084
Range	1.126	1.190
Classes/Bins	100	100
Class Width	0.011	0.012
Count	10000	10000

Figure 1.8: Minimum, Maximum, Range, Class Bins and Width of both ratios.

Frequency Distributions:				
DAM 1:	Class left	Class right	Class midpoint	Class Frequency
	0.918	0.930	0.924	1
	0.930	0.941	0.935	0
	0.941	0.952	0.946	0
	0.952	0.963	0.958	1
	0.963	0.975	0.969	1
	0.975	0.986	0.980	1

Figure 1.9: Tabular Frequency Distribution of Dam #1.

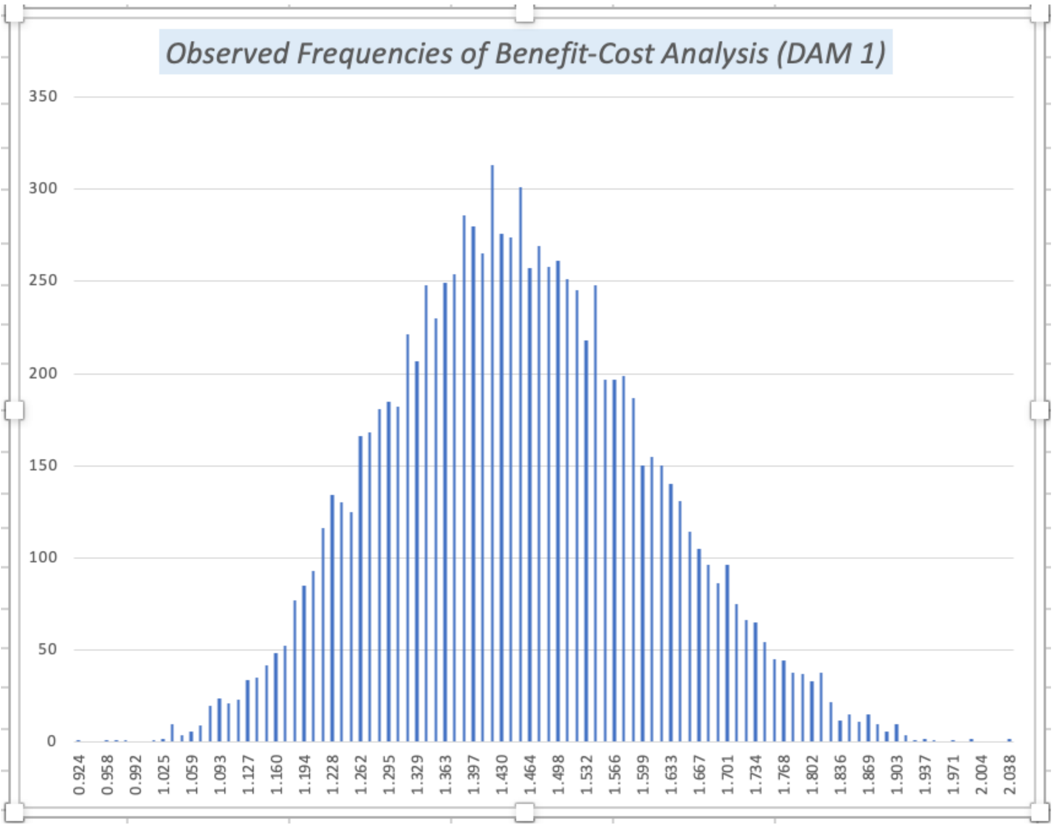


Figure 1.10: Graphical Observed Frequencies of Benefit-Cost Analysis (Dam #1).

The graphical representation of the frequencies (observed) of the Benefit-Cost ratios of Dam - #1 ***seems to be either normally distributed or gamma-distributed***. But, we cannot be sure of this and we would need to prove its distribution function using chi-squared test. We can overrule ***Beta distribution*** because the range of the values is more than range of Zero (0) to One (1).

The minimum, maximum, and range statistics are calculated from the Benefit-Cost ratios. Class bins and width are also calculated using these ratios which are used in itself for the "Class" distributions in namely "Class Left", "Class Right", "Class Midpoint", "Class Frequency".

	DAM 1 (α1)	DAM 2 (α2)
Min	0.918	0.894
Max	2.044	2.084
Range	1.126	1.190
Classes/Bins	100	100
Class Width	0.011	0.012
Count	10000	10000

Figure 1.11: Minimum, Maximum, Range, Class Bins and Width of both ratios.

DAM 2:	Class left	Class right	Class midpoint	Class Frequency
	0.918	0.930	0.924	2
	0.930	0.942	0.936	0
	0.942	0.954	0.948	1
	0.954	0.966	0.960	2
	0.966	0.978	0.972	3
	0.978	0.990	0.984	2

Figure 1.12: Tabular Frequency Distribution of Dam #2.

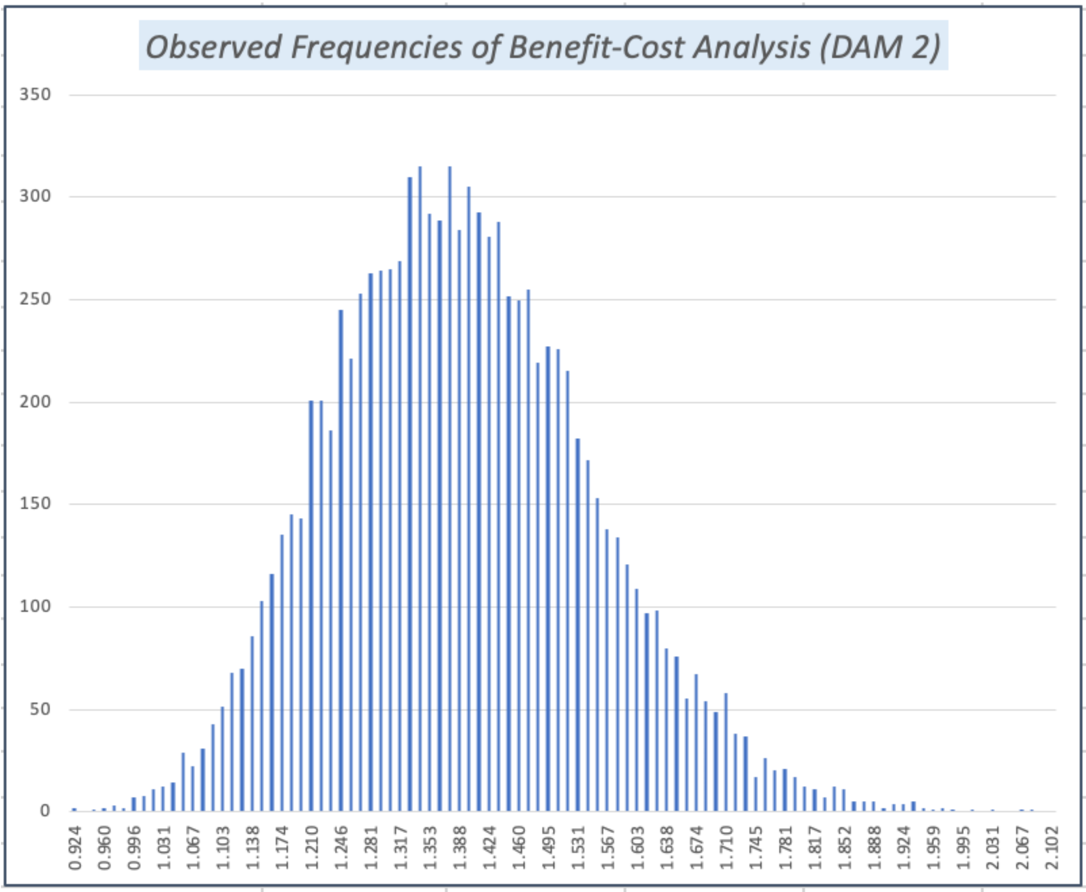


Figure 1.13: Graphical Observed Frequencies of Benefit-Cost Analysis (Dam #2).

The graphical representation of frequencies (observed) of the Benefit-Cost ratios of Dam - #2 seems to ***be either normally distributed or gamma-distributed***. But, we cannot be sure of this and we would need to prove its distribution function using chi-squared test. We can overrule ***Beta distribution*** because the range of the values is more than range of Zero (0) to One (1).

The minimum, maximum, and range statistics are calculated from the Benefit-Cost ratios. Class bins and width are also calculated using these ratios which are used in itself for the "Class" distributions in namely "Class Left", "Class Right", "Class Midpoint", "Class Frequency".

III) Some necessary calculations were performed using the values

Dam 1	Observed	Theoretical
Mean of the Total Benefits	29.620	29.600
SD of the Total Benefits	2.472	2.481
Mean of the Total Cost	20.535	20.800
SD of the Total Cost	1.539	1.562
Mean of the Benefit-cost Ratio	1.450	X
SD of the Benefit-cost Ratio	0.158	X
Dam 2	**Observed**	**Theoretical**
Mean of the Total Benefits	30.470	30.467
SD of the Total Benefits	2.386	2.381
Mean of the Total Cost	22.046	22.033
SD of the Total Cost	1.806	1.807
Mean of the Benefit-cost Ratio	1.391	X
SD of the Benefit-cost Ratio	0.157	X

Figure 1.14: Calculations were performed for the Mean, Standard Deviation (SD) of the Total Benefits & Total Cost, and Benefit-Cost Ratio of both Dams #1 and #2.

1. Observed Mean and Theoretical Mean of the total benefits of Dam#1 come out to be 29.620 and 29.6000. Respective values of Dam - #2 are : 30.470 & 30.467.

2. The observed and the theoretical Standard Deviation of the total involved benefits of Dam#1 come out to be 2.472 and 2.481. Respective values of Dam - #2 are : 2.386 & 2.381.

3. Observed Mean, and the theoretical Mean of the total costs of Dam - #1 come out to be 20.535 and 20.800. Respective values of Dam - #2 are : 22.046 & 22.033.

4. The observed and the theoretical Standard Deviation of the total costs of Dam #1 come out to be 1.539 and 1.562. Respective values of Dam #2 are 1.806 and 1.807.

5. For Benefit-Cost ratio of the construction project Dam #1, observed mean is 1.450 and observed standard deviation is 0.158.

6. For Benefit-Cost ratio of the construction project Dam #2, observed mean is 1.391 and observed standard deviation is 0.157.

7. These values are summation of theoretical values found earlier in part (i).

Part 2

Observing which theoretical probability distribution is good fit for the distribution of Dam #1

PART 2:	
Goodness of fit test	
Dam 1: Benefit-cost Ratio	
Mean ($\alpha 1$)	1.450
Variance ($\alpha 1$)	0.025
Alpha	83.953
Beta	0.017

Figure 2.1: Goodness of Fit Test statistics.

Dam 1:	Class left	Class right	Class midpoint	Class Frequency	Theotetical probability	Expected frequency	(Expected - Observed)^2 / Expected
	0.918	0.930	0.924	1	0.000	0.9	0.021
	0.930	0.941	0.935	0	0.000	0.4	0.398
	0.941	0.952	0.946	0	0.000	0.6	0.559
	0.952	0.963	0.958	1	0.000	0.8	0.064
	0.963	0.975	0.969	1	0.000	1.1	0.004
	0.975	0.986	0.980	1	0.000	1.4	0.139
	0.986	0.997	0.992	0	0.000	1.9	1.947
	0.997	1.008	1.003	0	0.000	2.6	2.587
	1.008	1.020	1.014	1	0.000	3.4	1.697

Figure 2.2: Theoretical, Expected Frequencies of Dam #1.

Chi-squared Test Statistic:	100.036
Chi-squared P-value:	0.396

Conclusion:

The P-value comes out to be greater than 0.05 which means we cannot reject the null hypothesis. We can say that there is not enough evidence for a difference present statistically between the distribution of α1 and its theoretical Gamma-distribution. It signifies that the Gamma-distribution is a good fit for distribution of α1 with 95% level of confidence.

Figure 2.3: Chi-Squared Test statistic, and P-value are calculated.

We can figure out the type of distribution of the distribution of Dam #1 (alpha 1) using our Chi-Squared Goodness-of-fit test -

1. The Chi-Squared Test statistics is the summation of the chi-squared metric values of each record ranging to the class' bins.

2. **We have used different types of distributions (examples : Beta, Normal, etc.) application in it where the NULL Hypothesis got rejected & we couldn't confirm or verify the type of this distribution. But, when applied Gamma distribution, we found significant results**.

3. **The Chi-Squared Test statistic comes out to be 100.036.**

4. Using this final Chi-Squared Metrics & Degree of Freedom, P-value is calculated of the distribution using an Excel function **CHISQ.TEST()** function.

5. The NULL Hypothesis,
   H_0: Gamma-distributions is an ideal & best kind of fit for this frequency distributions

6. The ALTERNATE Hypothesis,
   H_1: Gamma-distributions is NOT an ideal & best kind of fit for this frequency distributions

7. Alpha value = 0.05 (95% confidence). To verify the P-value, the series of theoretical and observed frequencies are used in the function **CHISQ.DIST().**

8. The P-value of the series comes out to be **0.396**. This P-value changes every time whenever random values are changed at each iteration. But, for this iteration, it is 0.396.

9. The P-values emerges out to be greater or more noteworthy than 0.05 which means we cannot reject the aforementioned null hypothesis. We are able to say that there are no sufficient evidences for a difference present statistically between the distributions of $\alpha 1$ and its theoretical Gamma distributions. It signifies that the Gamma distributions is an ideal and good fit for distributions of $\alpha 1$ with a 95% level of confidence.

Part 3

Observing the Benefit-Cost Ratios of Both the construction projects Dam #1 and Dam #2.

Dam 1	Dam 2	
Benefit-cost Ratio	Benefit-cost Ratio	α1-α2
1.437	1.484	-0.046
1.612	1.389	0.224
1.346	1.327	0.019
1.423	1.368	0.055
1.471	1.560	-0.088
1.316	1.601	-0.286
1.534	1.374	0.160
1.550	1.441	0.109
1.267	1.731	-0.465

Figure 3.1: Benefit-Cost Ratios of Both Dams #1 and #2 along with their differences.

	α_1	α_2
Minimum	0.999	0.957
Maximum	2.031	2.124
Mean	1.450	1.391
Median	1.442	1.382
Variance	0.025	0.025
Standard Deviation	0.157	0.158
SKEWNESS	0.254	0.321
$P(\alpha_i > 2)$	0.0004	0.0002
$P(\alpha_i > 1.7)$	0.0639	0.0363
$P(\alpha_i > 1.5)$	0.3623	0.2367
$P(\alpha_i > 1.3)$	0.8255	0.7029
$P(\alpha_i > 1)$	0.9999	0.9979

Figure 3.2: Theoretical, Expected Frequencies of Dam #1.

$P(\alpha_1 > \alpha_2)$	**0.604**

Figure 3.3: Probability difference.

1. The minimum value from the distributions of Benefit-Cost ratios of Dam#1 is 0.972 and of Dam#2 is 0.926.

2. The maximum value from the distributions of Benefit-Cost ratios of Dam#1 is 2.044 and of Dam#2 is 2.114.

3. The mean value of the distributions of the Benefit-Cost ratios of Dam -#1 is 1.445 & of Dam - #2 is 1.383.

4. P-values of Dam - #1 (α1) are usually greater/higher than the P-values of Dam-#2(α2)

5. The construction project Dam - #1 in the Southwest Georgia area should be picked up or considered instead of the construction project Dam#2 in North Carolina because of the higher or larger Benefit-Cost Ratio of Dam #1 over Dam #2.

CONCLUSION

We've performed Benefit-Cost Analysis of 2 construction projects (Dam #1 and Dam #2) available to a corporation. In this analysis, the estimations are made in different categories for both the annualized benefits & the annualized costs, deriving from/or belonging to projects of a corporation.

- The Triangular distribution is used to initialize the analysis with calculations of Theoretical and Observed values (experimental).
- We performed Monte Carlo simulations to retrieve different outcomes of all the benefits and costs (in millions of dollars) of both the construction projects (Dam - #1 and Dam - #2).
- The Benefit-Cost Ratios distribution of both of the projects were calculated out and used to figure out the type of distributions in them.
- The Chi-Squared Goodness-of-fit test was applied to figure out the type of distribution of the 1st project (Dam #1). After various tries, **we found that the Gamma Distribution is found to be the best and ideal fit for the presented distribution**.
- **We later analysed that the P-values of the construction project Dam #1 ($\alpha 1$) is usually greater or larger than the P-values of the construction project Dam #2 ($\alpha 2$).**
- This signifies that the JET Corporation should pick up the construction project of dam based in Southwest Georgia area (Dam #1) over the second construction project of dam based in North Carolina area because of the higher or larger Benefit-Cost Ratio it provides to the company and greater P-values in most of the cases.

BIBLIOGRAPHY

1. Microsoft. (2021). *Excel functions (alphabetical)*. https://support.microsoft.com/en-us/office/excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188

2. *Triangular Distribution*. (2021, June 29). Real Statistics Using Excel. https://www.real-statistics.com/other-key-distributions/uniform-distribution/triangular-distribution/

3. Z. (2021e, March 28). *An Introduction to the Triangular Distribution*. Statology. https://www.statology.org/triangular-distribution/

4. ALY 6050 - Prof Richard He Lab - *The Triangular Probability Distributions & Random Number Generation* (2022, January), https://northeastern.instructure.com/courses/97784/pages/module-2-lab-2-the-triangular-probability-distributions-and-random-number-generation?module_item_id=6650722