Benefit-Cost Ratio analysis in construction projects

Heejae Roh
Northeastern University
ALY 6050: Intro to Enterprise Analytics
Azadeh Mobasher
March 5, 2023

PART 0. INTRODUCTION

In this module, I will run a simulation on corporation selection among two projects. Based on the benefit and cost data, I will run 10,000 simulations to determine which option 1 or option 2 is better with benefit-cost analysis. In the process, I will run a chi-square test to see what the distribution we created looks like. And finally, based on the generated data, I will analyze the result value and make the final choice. This is a problem in which a company chooses between options 1 and 2 when building a dam. I will analyze the problem using 6 benefit factors and 2 cost factors.

PART 1. ANALYSIS

Understanding Key metrics

1. All benefit and cost is in the form of a triangular distribution
2. Benefits (6): improved navigation, hydroelectric power, fish and wildlife, recreation, flood control, and the commercial development of the area
3. Three estimates available for each type benefit – a minimum possible value, a most likely value (i.e., a mode or peak), and a maximum possible value
4. Costs (2): two categories associated with a construction project of this type have been identified: the total capital cost, annualized over 30 years (at a rate specified by the creditors and the government), and the annual operations and maintenance costs
5. These benefits and costs estimations for both dam projects (in millions of dollars) are as follows:

Table 1: Benefits and costs for the Dam #1 construction project in millions of dollars

Dam #1: Benefits & Costs

Benefit	Estimate		
	Minimum	*Mode*	*Maximum*
Improved navigation **B1**	1.1	2	2.8
Hydroelectric power **B2**	8	12	14.9
Fish and wildlife **B3**	1.4	1.4	2.2
Recreation **B4**	6.5	9.8	14.6
Flood control **B5**	1.7	2.4	3.6
Commercial development **B6**	0	1.6	2.4
Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	13.2	14.2	19.1
Operations & Maintenance **C2**	3.5	4.9	7.4

Table 2: Benefits and costs for the Dam #2 construction project in millions of dollars

Dam # 2: Benefits & Costs

Benefit	Estimate		
	Minimum	*Mode*	*Maximum*
Improved navigation **B1**	2.1	3	4.8
Hydroelectric power **B2**	8.7	12.2	13.6
Fish and wildlife **B3**	2.3	3	3
Recreation **B4**	5.9	8.7	15
Flood control **B5**	0	3.4	3.4
Commercial development **B6**	0	1.2	1.8
Cost	*Minimum*	*Mode*	*Maximum*
Annualized capital cost **C1**	12.8	15.8	20.1
Operations & Maintenance **C2**	3.8	5.7	8

PART 1. Simulation & Parameters & Distribution

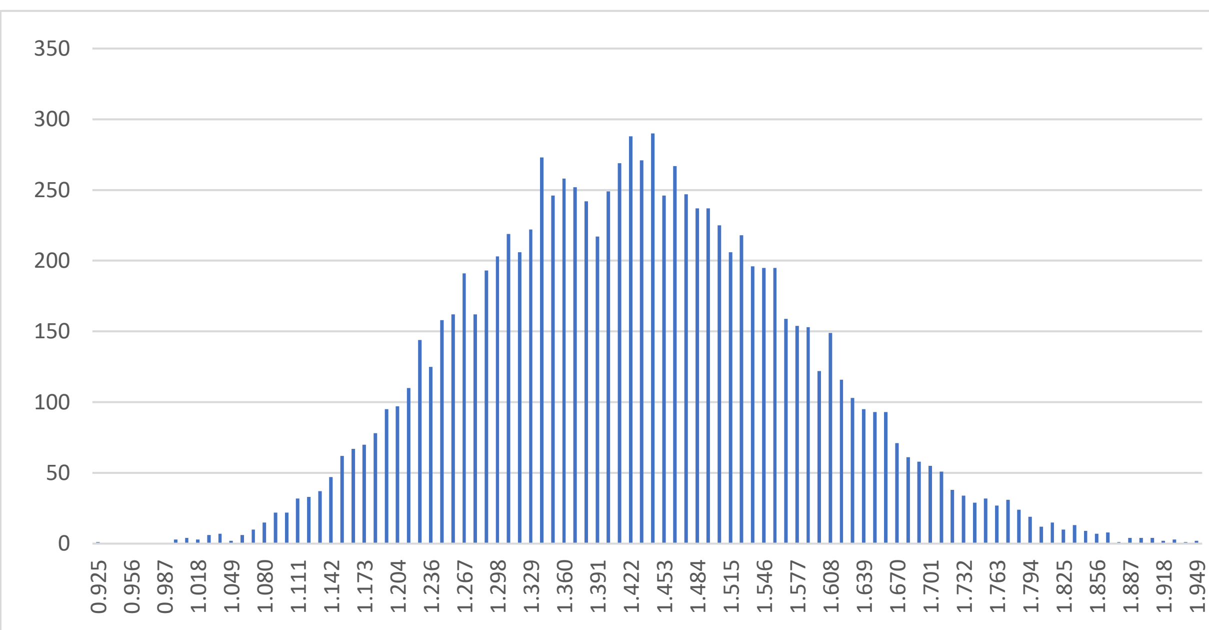
Dam #1 Benefit simulation head five of 10,000

B1	B2	B3	B4	B5	B6	SUM[1]	C1	C2	SUM[2]	Benefit -Cost Ratio
1.79	12.23	1.59	8.73	2.76	0.65	27.75	14.34	4.35	18.69	1.617
2.34	11.68	1.58	7.06	2.53	1.52	26.71	14.48	6.00	20.48	1.473
2.46	10.63	1.73	12.08	2.08	1.61	30.59	16.51	5.72	22.23	1.418
2.29	10.11	1.48	11.54	2.43	1.10	28.94	18.12	4.30	22.42	1.433
1.65	12.38	2.05	11.85	2.25	0.85	31.03	14.91	4.52	19.43	1.343
		…			…					

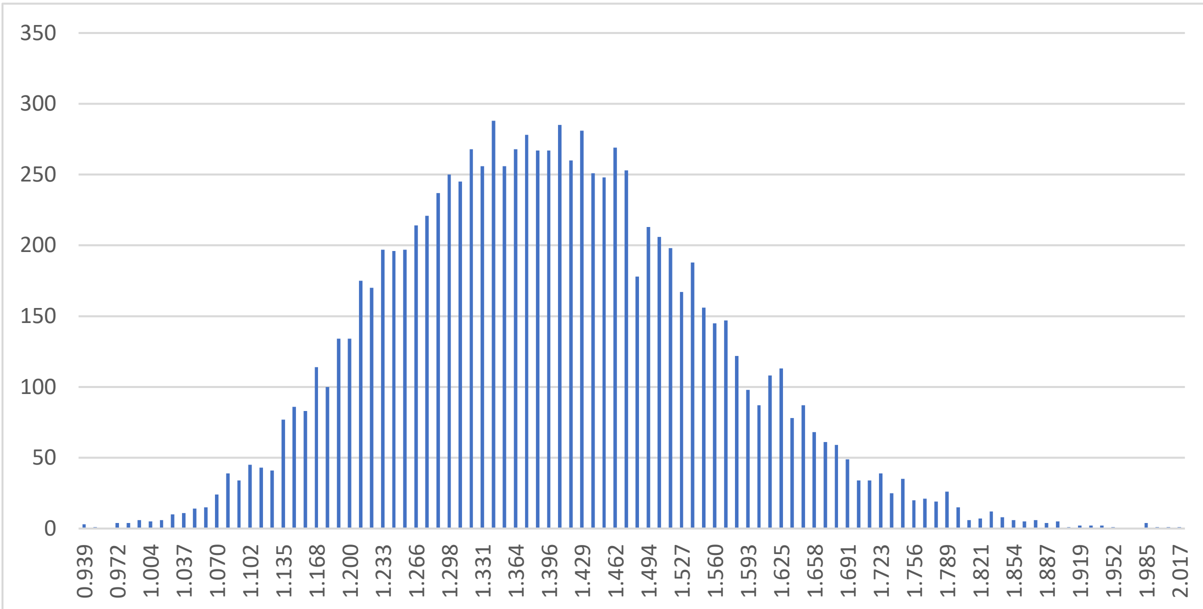
10,000 Random Values

1. Using RAND(), Create 10,000 random values
2. With triangular distribution analysis methods, a= left end, b= right end, c= peak, K=(c-a)/ (b-a), A=(b-a)(c-a), B=(b-a)(b-c)
3. Using if r <= K, then x= a+ SQRT(r*a), if r > K, then b- SQRT(1-r)*B, calculate the values of triangular distribution
4. SUM[1] is the total of benefits from Dam #1 project, and SUM[2] is the total of benefits from Dam #1 project
5. Benefit-cost Ratio is calculated by SUM[1] / SUM[2]. If the ratio is greater than 1, this means that benefits are greater than costs

Benefit/ Cost Ratio of Distribution of Dam #1



Benefit/ Cost Ratio of Distribution of Dam #2



Interpretation

1. This is the Benefit/Cost Ratio (BCR) in a histogram
2. Mean of Benefit/Cost Ratio is 1.43 for Dam #1, slightly larger than Dam #2's 1.40.
3. SD (Standard Deviation) of Benefit/Cost Ratio is 0.15 for Dam #1, slightly smaller than 0.16 for Dam #2
4. Dam #1 has a higher value on average. And has lower SD which means that it has more values closer to the mean
5. In the graph, Dam #1 has a narrower range, but Dam #2 has a wider range

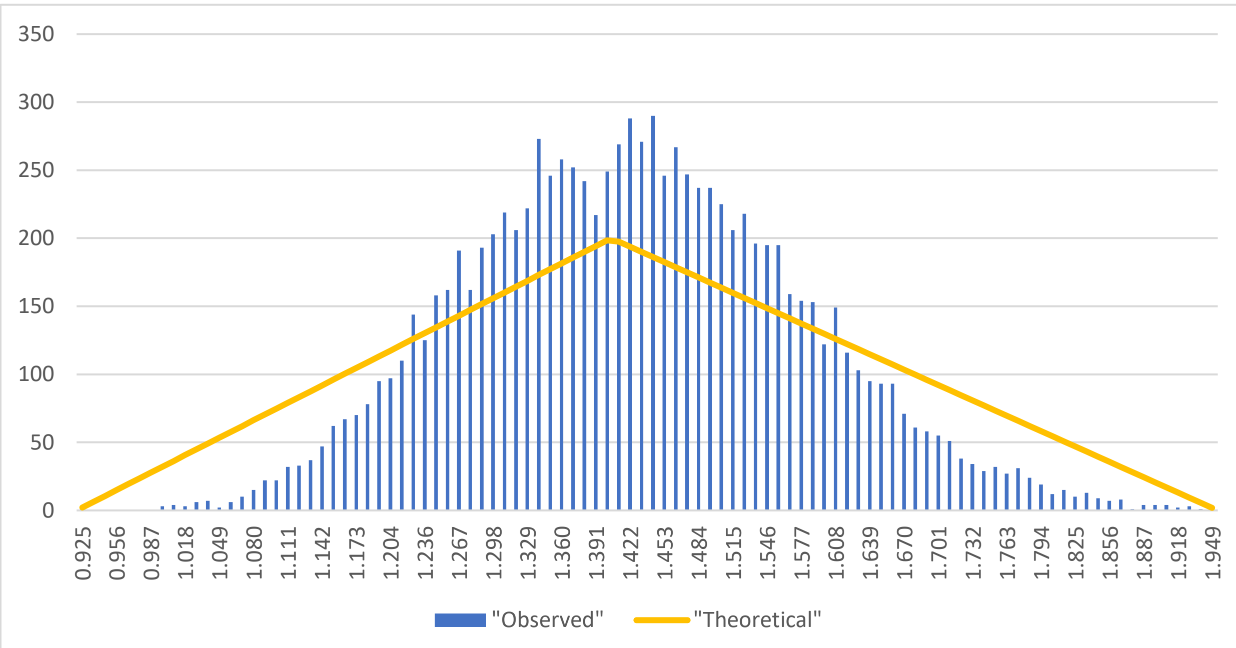
Table: Parameters for Benefits & Costs for each

	Dam #1		Dam #2	
	Observed	Theoretical	Observed	Theoretical
Mean of the Total Benefits	29.50	29.50	30.71	30.70
SD of the Total Benefits	2.31	2.31	2.42	2.41
Mean of the Total Cost	20.76	20.76	22.07	22.07
SD of the Total Cost	1.53	1.53	1.74	1.73
Mean of the Benefit-cost Ratio	1.43	X	1.40	X
SD of the Benefit-cost Ratio	0.15	X	0.16	X

Interpretation

1. Mean of Benefit in Dam #2 was 1.21 greater than Dam #1's. Mean of the total benefits are 29.50 and 30.71 for Dam #1/ Dam #2 respectively
2. SD of Benefit and Cost in Dam #1 was smaller than Dam #2. SD of the Benefit and Cost of Dam #1 is 2.31 and 1.53, and SD of the Benefit and Cost of Dam #2 is 2.42 and 1.74
3. This means that the Benefit and Cost of Dam #1 is closer to the Mean comparing with Dam #2

Comment on Distribution 1



1. I cannot not find any basis for constructing a new triangular distribution by simply adding the left end, right end, and peak of each triangular distribution
2. However, six Benefit and two Cost are from a triangular distribution. Let's assume a triangular distribution first and perform the chi-square test in PART 2
3. After drawing a triangle in the graph above doesn't seem to fit well
4. Let's find out exactly through the test

PART 2-1. Goodness of fit test for triangular distribution

Understanding

1. A triangular distribution is a continuous probability distribution with a probability density function shaped like a triangle. It is defined by three values: the minimum value a, the maximum value b, and the peak value c (statslc, n.d.)
2. The triangular distribution provides a simplistic representation of the probability distribution when limited sample data is available. Its parameters are the minimum, maximum, and peak of the data (MathWork, 2022)

Triangular Distribution:

left end - **a**	peak - **c**	right end - **b**
0.955	1.317	2.013

c-a	**b-a**	**b-c**
0.362	1.058	0.696

K = (c-a) / (b-a)	**A =(b-a) * (c-a)**	**B = (b-a) * (b-c)**
0.342	0.383	0.737

Interpretation

1. Left end (a) is calculated by the Minimum function with 10,000 simulation results
2. right end (b) is calculated by the Maximum function with 10,000 simulation results
3. peak (c) is calculated by the formular ‘3*μ – a – b’ which is derived from triangular mean formular

Head five of calculating chi-square statistics

Class left	Class right	Class midpoint	Class Frequency	Theotetical probability	Expected frequency	(Expected - Observed)^2 / Expected
0.955	0.965	0.96	2	0.0003	2.923	0.292
0.965	0.976	0.971	1	0.0009	8.77	6.884
0.976	0.987	0.981	0	0.0015	14.617	14.617
0.987	0.997	0.992	2	0.002	20.463	16.659
0.997	1.008	1.003	0	0.0026	26.31	26.31
					SUM	2327.25

Chi-squared goodness of Fit Test

Step 0 Finding Key-Metrics

This test is to check which distribution the simulation we created follows. Check for triangular distribution

Step 1 Hypothesis.

Null: The Triangular distribution is a good fit for this distribution

Alternative: The Triangular distribution is not a good fit for this distribution

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

Chi-squared Test Statistic:	2853.35
Chi-squared P-value:	0.000
DF	96

- The p-value with 1-CHISQ.DIST() in excel is 0.00

Step 4 Make the decision.

- There is enough evidence to reject null hypothesis
  since 0.00 (p-value) < 0.05

Step 5 Summarize the results.

- Don’t accept that the triangular distribution is a good fit

Interpretation

1. Chi-squared test statistic is SUM of 100 $(\text{Expected} - \text{oberseved})^2 / \text{Expected}$
2. Df = Number of observations – Number of parameters estimated -1
3. Number of observations are 100, number of parameters estimated are 3 which are a, b, c. Therefore, DF is 100 minus 3 minus 1 is 96
4. P-value calculated by 1-CHISQ.DIST() is 0.00

PART 2-2. Goodness of fit test for GAMMA distribution

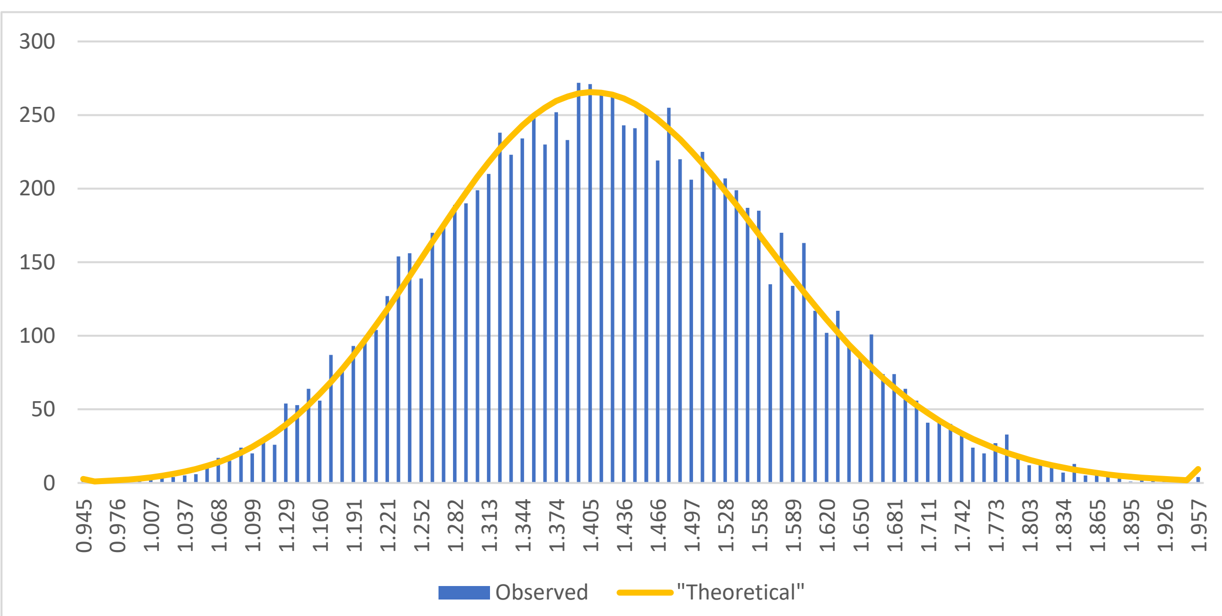
Understanding

1. Gamma distributions are common in engineering models. For example, time to failure of equipment and load levels for telecommunication services, meteorology rainfall, and business insurance claims and loan defaults, where the variables are always positive and the results are skewed (Divya, 2022)
2. Suppose X contains sampled historical data indexed by I. To estimate the parameters of the gamma distribution that best fits this sampled data, the following parameter estimation formulae can be used (analytica, n.d.):
   alpha: Mean(X, I)^2/Variance(X, I)
   beta: Variance(X, I)/Mean(X, I)
3. Alpha = Mean^2/Variance = 1.427^2/0.023 = 88.643
   Beta = Variance/Mean = 0.023/1.427 = 0.016

Head five of calculating chi-square statistics

Class left	Class right	Class midpoint	Class Frequency	Theotetical probability	Expected frequency	(Expected - Observed)^2 / Expected
0.897	0.908	0.903	1	0	0.438	0.722
0.908	0.919	0.914	0	0	0.217	0.217
0.919	0.93	0.925	0	0	0.313	0.313
0.93	0.941	0.936	0	0	0.445	0.445
0.941	0.952	0.947	0	0.0001	0.625	0.625
					SUM	102.38

Gamma distribution



Chi-squared goodness of Fit Test

Step 0 Finding Key-Metrics
This test is to check which distribution the simulation we created follows. Check for gamma distribution.

Step 1 Hypothesis.
Null: The Gamma distribution is a good fit for this distribution
Alternative: The Gamma distribution is not a good fit for this distribution

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

Chi-squared Test Statistic:	102.380
Chi-squared P-value:	0.335
DF	97

- The p-value with 1-CHISQ.DIST() in excel is 0.335

Step 4 Make the decision.

- There is not enough evidence to reject null hypothesis since 0.335 (p-value) > 0.05

Step 5 Summarize the results.

- Accept that the Gamma distribution is a good fit

Interpretation

1. Chi-squared test statistic is SUM of 100 (Expected – oberseved)^2/ Expected. Which is 102.38
2. Number of observations are 100, number of parameters estimated are 2 which are alpha, and beta. Therefore, DF is 100 minus 2 minus 1 is 97
3. P-value calculated by 1-CHISQ.DIST() is 0.335

PART 3. Final decision of projects

	$\alpha 1$	$\alpha 2$
Minimum	0.950	0.933
Maximum	2.032	2.040
Mean	1.426	1.401
Median	1.423	1.394
Variance	0.023	0.024
Standard Deviation	0.152	0.154
SKEWNESS	0.154	0.320
$P(\alpha i > 2)$	0.0002	0.0001
$P(\alpha i > 1.8)$	0.0074	0.0097
$P(\alpha i > 1.5)$	0.3114	0.2499
$P(\alpha i > 1.2)$	0.9339	0.9091
$P(\alpha i > 1)$	0.9993	0.9987

Interpretation

1. On mean, alpha 1 is 0.025 higher than alpha 2. This indicates that Dam #1 outperforms the BCR a bit more
2. As for Variance, alpha 1 is smaller than alpha 2. It means that the Dam #1 project can get more similar result
3. Comparing the maximum and minimum values of alpha 1 and alpha 2, the minimum value is larger in alpha 1, and the maximum value is smaller in alpha 1. In other words, the range of Dam #1 is narrower and Dam #2 is wider
4. The probability that alpha 1 is greater than 1 is 0.9993, and the probability that alpha 2 is greater than 1 is 0.9987
5. Mainly, if the probability the alpha is greater than 1, it can be said that the benefit is higher than the cost in this analysis, which is an important number. Alpha 1 has higher possibility to be greater than 1, is a more stable business comparing with Dam #2
6. Probability that the alpha greater than 1.5 is 0.3114 at alpha 1, and 0.2499 at alpha 2. I can expect to record a larger BCR at alpha 1
7. The skewness is positive in both alpha 1 and alpha 2, it is more likely to similar with Gamma distribution

Comparing Dam #1 and Dam #2	
$P(\alpha_1 > \alpha_2)$	0.554

1. Finally, comparing each random alpha 1 and alpha 2, the possibility of alpha 1 is greater than alpha 2 is 0.554
2. It means if we repeats 100 projects, we can get more BCR in 55 project in Dam #1 comparing with Dam #2
3. The decision is choosing Dam #1

4. And, assuming that the company wants a more aggressive investment, I check the alpha 2 project again, but the advantage over the alpha 1 project was only when it was greater than 1.8, and this probability was also less than 1%. Therefore, I would recommend the Dam #1 project more strongly

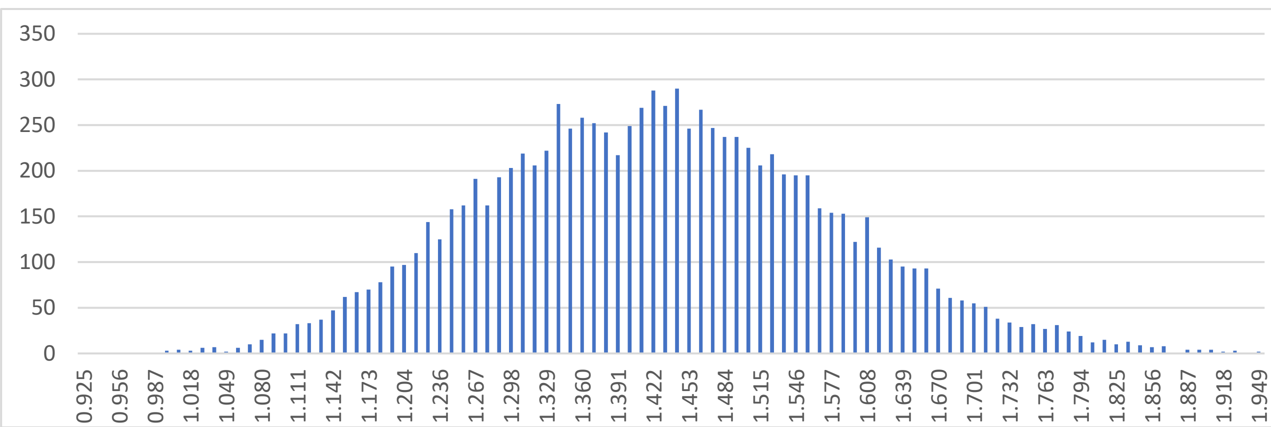
PART 4. Answers for Questions

PART 1

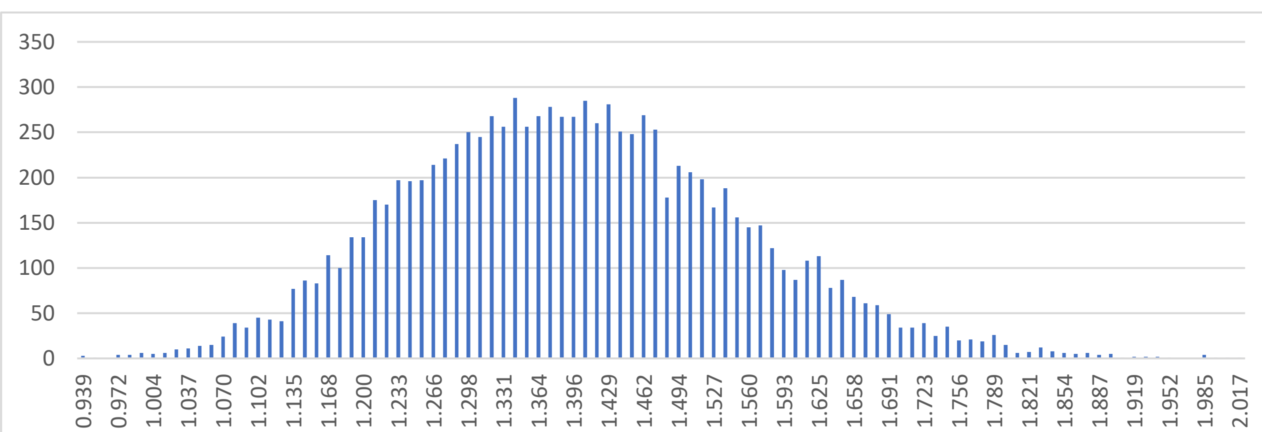
1. <u>Perform a simulation of 10,000 benefit-cost ratios for Dam #1 & Dam #2</u>
   Performed the simulation of 10,000 benefit-cost ratios (BCR) for Dam #1 & Dam #2. There were 6 benefit factors and 2 cost factors for each project. All factor is based on triangular distribution which has 3 parameters, left end (a), right end (b), and peak (c). Used RAND () function in Excel, and calculate the exact values from 3 parameters in 10,000 random numbers

2. <u>Construct a both a tabular and a graphic frequency distribution. Include only the graphical distributions and comment on the shape of each distribution</u>

   Benefit – Cost Ratio of Distribution of Dam #1



   Benefit – Cost Ratio of Distribution of Dam #2



3. Calculating the Observed & Theoretical parameters for Dam #1 & Dam #2

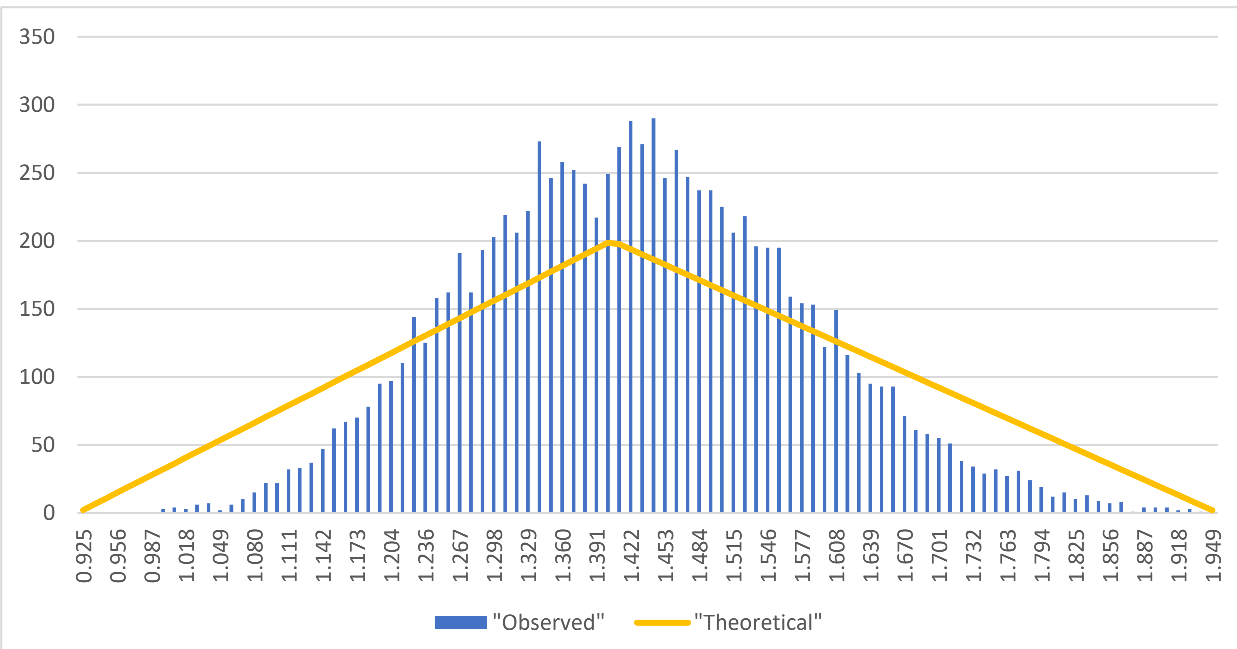
Table: Parameters for Benefits & Costs for each

	Dam #1		Dam #2	
	Observed	Theoretical	Observed	Theoretical
Mean of the Total Benefits	29.50	29.50	30.71	30.70
SD of the Total Benefits	2.31	2.31	2.42	2.41
Mean of the Total Cost	20.76	20.76	22.07	22.07
SD of the Total Cost	1.53	1.53	1.74	1.73
Mean of the Benefit-cost Ratio	1.43	X	1.40	X
SD of the Benefit-cost Ratio	0.15	X	0.16	X

PART 2

1. Select a theoretical probability distribution that is good fit for your distribution. Describe the rational for your choice of the probability distribution and a description of the outcomes. Indicate the values of the chi-squared test statistic and the P-value

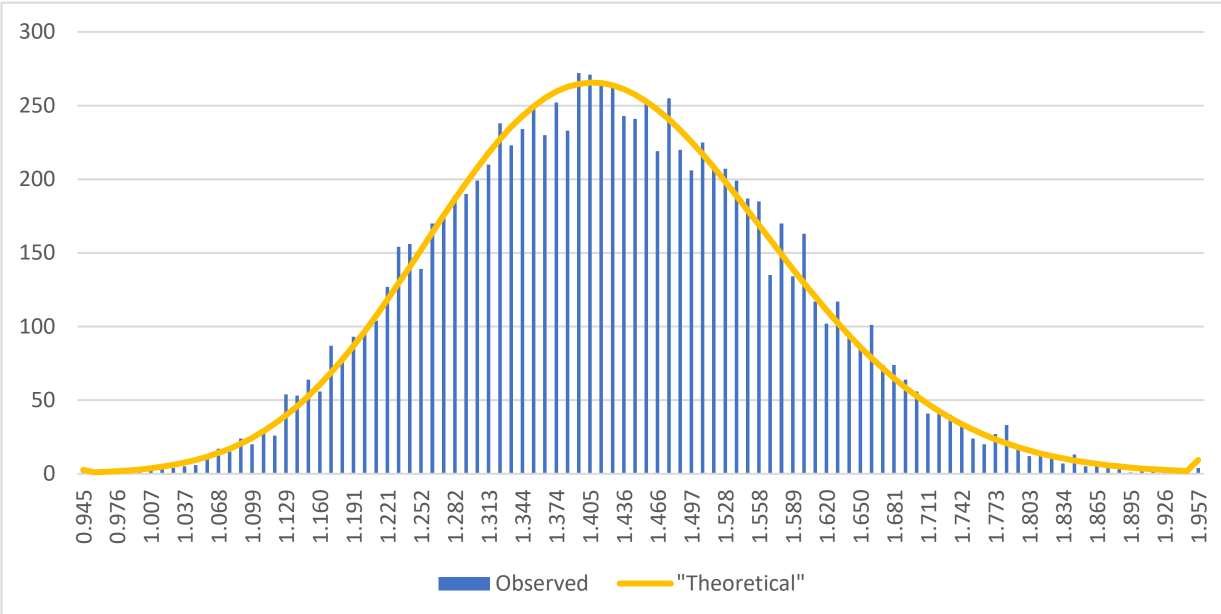
Triangular distribution



First, selected the triangular distribution for my random distribution. The reason is that every benefit and cost is made by triangular distribution, so assume a triangular distribution and perform the chi-squared test

However, the P-value from the chi-squared test is 0.000 which is less than 0.05. The decision is ‘Don’t accept that the triangular distribution is a good fit

Gamma distribution



After chi-squared test with triangular distribution, I plan test with gamma distribution test, because it looks positively skewed and not Gaussian distribution. I calculated the alpha and beta for gamma distribution. With alpha and beta, made a formular to calculated the theoretical values for each range.
The p-value was 0.335, there is not enough evidence to reject the null-hypothesis that the Gamma distribution is a good fit.

PART 3

1. Complete the table about project of Dam #1 and Dam #2

	$\alpha 1$	$\alpha 2$
Minimum	0.950	0.933
Maximum	2.032	2.040
Mean	1.426	1.401
Median	1.423	1.394
Variance	0.023	0.024
Standard Deviation	0.152	0.154
SKEWNESS	0.154	0.320
$P(\alpha i > 2)$	0.0002	0.0001
$P(\alpha i > 1.8)$	0.0074	0.0097
$P(\alpha i > 1.5)$	0.3114	0.2499
$P(\alpha i > 1.2)$	0.9339	0.9091
$P(\alpha i > 1)$	0.9993	0.9987

2. Recommend one of two projects to the management. Explain all your rationales for the project. Include with the final conclusion of your report an estimate for the probability that alpha 1 will be greater than alpha 2

Comparing Dam #1 and Dam #2	
$P(\alpha_1 > \alpha_2)$	0.554

The decision is choosing Dam #1 rather than Dam #2. The possibility of alpha 1 is greater than alpha 2 is 0.554. Assuming that the company wants a more aggressive investment, I check the alpha 2 project again, but the advantage over the alpha 1 project was only when it was greater than 1.8, and this probability was also less than 1%. And since the mean is larger, the median is larger, and the variance is smaller, alpha 1 will benefit the company more than alpha 2

PART 5. CONCLUSION

In this module, we solved the benefit-cost ratio problem through simulation. I learned how to use the triangular distribution in practice. It is often used in simulations when there is very little known about the data-generating process and is often referred to as a “lack of knowledge” distribution (Aurel, 2013). In my opinion, if there is a lack of information in a real business situation, I can use triangular distribution. Or, when information is lacking, I think I can approach triangular distribution first.

And while doing the goodness fit test, I thought of all the other distributions. The distributions were Uniform / Triangular / Exponential / Normal / Standard Beta / Gamma / Log-Normal / Weibull / Chi-Squared, which I learned last time. As a result of searching for each and examining the characteristics, I thought that the gamma distribution was the most similar to my randomly generated distribution, so I proceeded the test with GAMMA. In the process, I learned how to match random numbers based on each distribution.

REFERENCE

Svetlana, Cheusheva. (February 7, 2023). How to calculate median in Excel - formula examples. Ablegits.com. Retrieved from https://www.ablebits.com/office-addins-blog/median-formula-excel/

Microsoft. (n.d.). SKEW function. Retrieved from https://support.microsoft.com/en-us/office/skew-function-bdf49d86-b1ef-4804-a046-28eaea69c9fa

statslc. (n.d.). Triangular distributions. Retrieved from https://learnandteachstatistics.files.wordpress.com/2013/07/notes-on-triangle-distributions.pdf

MathWorks. (2022). Triangular distributions. Retrieved from https://www.mathworks.com/help/stats/triangular-distribution.html

Divya, Kare. (September 6, 2022). Gamma distribution: definition, example, properties and applications. testbook. Retrieved from https://testbook.com/learn/maths-gamma-distribution/

analyticsa. (n.d.). Gamma distribution. Retrieved from https://docs.analytica.com/index.php/Gamma_distribution

Aurel, Spătaru. (2013). Analysis and probability. ScienceDirect. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780124016651000072