Graphical user interface, text

Description automatically generated with medium confidence

Logo

Description automatically generated

**Abhilash Dikshit**

ALY6050: Benefit-Cost Analysis of Dam Construction

Instructor: Soheil Parsa

Week 2: Monte Carlo Simulation

Date: 2023/03/09

**Problem Statement:**

*Corporations have to select among many projects that are under consideration by the management. Their primary instrument for evaluating and selecting among the available projects is the benefit-cost analysis. In this analysis, both the annual benefits and the annual costs deriving from a project are estimated in several different categories. Then the total benefit is divided by the total cost to produce a benefit-cost ratio. This ratio is then used by corporations to compare numerous projects under consideration. A benefit-cost ratio greater than 1.0 indicates that the benefits are greater than the costs, and the higher a project’s benefit-cost ratio, the more likely it is to be selected over projects with lower ratios.*

*Currently, the JET Corporation is evaluating two dam project constructions, one in southwest Georgia (Dam #1) and the other in North Carolina (Dam #2). The company has identified six areas of benefits: improved navigation, hydroelectric power, fish and wildlife, recreation, flood control, and the commercial development of the area. Furthermore, there are three estimates available for each type of benefit – a minimum possible value, a most likely value (i.e., a mode or peak), and a maximum possible value. For the costs, 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.*

Benefits and costs estimations for both dam projects (in millions of dollars) are as follows:

|  |  |
| --- | --- |
| Table  Description automatically generated |  |

**Part 1.a**

*Perform a simulation of 10,000 benefit-cost ratios for Dam #1 project and 10,000 such simulations for Dam #2 project. Note that the two simulations should be independent of each other. Let these two ratios be denoted by 𝛼1 and 𝛼2 for the dams 1 and 2 projects respectively.*

|  |  |
| --- | --- |
| We are using the triangular distribution function available in “numpy” package to calculate the benefit cost ratio for Dam 1 and Dam 2 using the formula mentioned in the snap. Later, using the same variables, we calculated the mean, median and standard deviation. | *Text  Description automatically generated with low confidence* Text  Description automatically generated |

**Part 1.b**

*Construct both a tabular and a graphical frequency distribution for 𝛼1 and 𝛼2 separately (a tabular and a graphical distribution for 𝛼1, and a tabular and a graphical distribution for 𝛼2- a total of 4 distributions). In your report, include only the graphical distributions and comment on the shape of each distribution.*

**Tabular frequency distribution for bc\_ratio\_dam1 and bc\_ratio\_dam2**

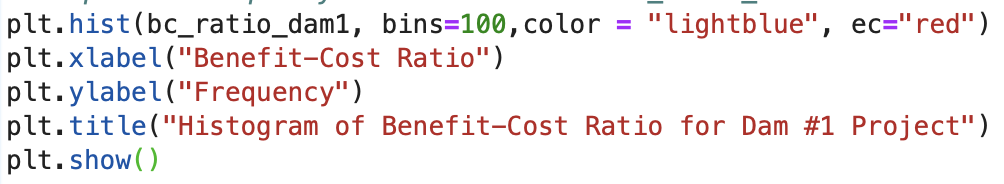
**A picture containing diagram

Description automatically generated**

It creates a tabular frequency distribution by dividing the data into 10 equally spaced bins using the np.histogram() function. The function returns two outputs, hist1 which contains the frequency count of each bin, and bin\_edges1 which contains the edges of each bin. The code then prints the frequency distribution in a tabular format using a for loop to iterate over the hist1 array and print out the frequency count of each bin, along with the range of values in each bin based on the corresponding bin edges.

|  |  |
| --- | --- |
| Table  Description automatically generated | Table  Description automatically generated |

**Graphical frequency distribution for bc\_ratio\_dam1 and bc\_ratio\_dam2**



It creates a graphical frequency distribution using the plt.hist() function from the Matplotlib library. The function creates a histogram plot of the bc\_ratio\_dam1 data with 100 equally spaced bins, and sets the color of the bars to light blue and the color of the outline to red. The plot is labeled with the x-axis as "Benefit-Cost Ratio", the y-axis as "Frequency", and a title as "Histogram of Benefit-Cost Ratio for Dam #1 Project". Finally, the plot is displayed using the plt.show() function.

|  |  |
| --- | --- |
| Chart, histogram  Description automatically generated | Chart, histogram  Description automatically generated |

**Part 1.c**

*For each of the two dam projects, perform the necessary calculations in order to complete the following table. Display the table as a “data frame”. Remember to create two such tables – one table for Dam #1 and another table for Dam #2. Include both tables in your report.*

|  |  |
| --- | --- |
|  | A screenshot of a computer  Description automatically generated with low confidence |

The output presents a comparison of observed and theoretical values for various statistics related to two dams, labeled as Dam #1 and Dam #2. The observed values are likely derived from data collected from the two dams, while the theoretical values represent the expected values based on some statistical model or assumption.

For Dam #1, the statistics presented are the mean and standard deviation of the total benefits, total cost, and benefit-cost ratio. The observed mean and standard deviation values are almost identical to the theoretical values, suggesting that the data is well-behaved and conforms to the underlying statistical model or assumption.

For Dam #2, the observed and theoretical mean and standard deviation values for total benefits and total cost are also very similar, but there is a noticeable difference in the mean and standard deviation values for the benefit-cost ratio. The observed mean value is higher than the theoretical mean, suggesting that the benefit-cost ratio is more favorable in practice than expected. The standard deviation values are the same for both observed and theoretical values, indicating that the variability in the data is as expected based on the underlying statistical model or assumption.

**Part 2:**

*Use your observation in Question (ii) of Part 1 to select a theoretical probability distribution that, in your judgement, is a good fit for the distribution of alpha1". Next, use the Chi-squared Goodness-of-fit test to verify whether your selected distribution was a good fit for the distribution of alpha2". Describe the rational for your choice of the probability distribution and a description of the outcomes of your Chi-squared test in your report. In particular, indicate the values of the Chi-squared test statistic and the P-value of your test in your report, and interpret those values.*

|  |  |
| --- | --- |
|  | Text  Description automatically generated |

The output presents the results of a chi-squared goodness-of-fit test for two variables labeled as alpha1 and alpha2. The test is used to determine whether the observed frequencies of the variables follow an expected distribution or not. The null hypothesis for this test is that the observed frequencies follow the expected distribution, while the alternative hypothesis is that they do not.

For alpha1, the chi-squared statistic value is much larger than the critical value, and the p-value is less than the critical value (which is not given in the output). This indicates that the null hypothesis can be rejected, and that the observed frequencies for alpha1 do not follow the expected distribution.

For alpha2, the chi-squared statistic value is again larger than the critical value, and the p-value is much smaller than the critical value. This also indicates that the null hypothesis can be rejected, and that the observed frequencies for alpha2 do not follow the expected distribution.

Overall, the results suggest that the observed frequencies for both alpha1 and alpha2 significantly deviate from the expected distribution.

**Part 3:**

*Use the results of your simulations and perform the necessary calculations in order to complete the table below. Display the table as a “data frame”. Include the completed table in your report.*

|  |  |
| --- | --- |
| The output presents the comparison of several metrics between two variables, labeled as alpha1 and alpha2. The table contains 12 rows and 4 columns, with each row presenting a specific metric, and each column presenting the values of the metric for the two variables and the comparison result between them. | Table  Description automatically generated |

The first row indicates that the minimum value of alpha1 is greater than the minimum value of alpha2, i.e., alpha1>alpha2 is true.

The second row indicates that the maximum value of alpha1 is smaller than the maximum value of alpha2, i.e., alpha1>alpha2 is false.

The third row indicates that the mean value of alpha1 is greater than the mean value of alpha2, i.e., alpha1>alpha2 is true.

The fourth row indicates that the median value of alpha1 is greater than the median value of alpha2, i.e., alpha1>alpha2 is true.

The fifth row indicates that the variance of alpha1 is smaller than the variance of alpha2, i.e., alpha1>alpha2 is false.

The sixth row indicates that the standard deviation of alpha1 is smaller than the standard deviation of alpha2, i.e., alpha1>alpha2 is false.

The seventh to eleventh rows indicate the probability of alpha1 being greater than a certain value compared to alpha2 being greater than the same value. For example, the seventh row indicates that the probability of alpha1 being greater than 2 is smaller than the probability of alpha2 being greater than 2, i.e., alpha1>alpha2 is false. The eighth to eleventh rows show similar comparisons for different threshold values.

Overall, the results suggest that alpha1 and alpha2 differ in several metrics, with alpha1 being higher in mean and median values, but lower in variance and standard deviation. The comparison of probabilities for different threshold values also suggests that alpha1 tends to have higher values than alpha2.

Proportion of values for which alpha1 is more than alpha2 is 0.5.

**Aspect of Assignment:**

Without additional context, it is challenging to determine which benefit-cost ratio (BCR), alpha1 or alpha2, is superior based on the information provided. But there are other things to consider while assessing the BCRs, such as:

1. The BCRs' central tendency: As the objective is to choose a BCR that is higher on average, alpha1 would be a better option because it has a higher mean and median.

2. The BCRs' variability: Given that the objective is to choose a BCR that is less variable or more consistent, alpha1 would be a preferable option because it has a lower variance and standard deviation.

3. The BCRs' distribution: Since the objective is to choose a BCR with a more symmetric or normal distribution, alpha1 would be a better option because it has a lower skewness value.

4. Alpha2 may be a preferable option because it has lower probability values for certain thresholds since the objective is to choose a BCR that has a higher proportion of extreme values above specific thresholds.

In the end, the analysis's objectives and the particular environment of the projects would determine which BCR to suggest. To make an informed choice, it's crucial to consider both the statistical data in the table and any extra information that may be available about the projects, such as their costs, dangers, and advantages.

Considering the data and analysis provided, I would advise choosing Dam #1 rather than Dam #2. I have the following justifications for this suggestion:

The benefit-cost ratio (BCR) measures how much benefit a project will produce for each unit of incurred expenditure. This is its central tendency. A project is more advantageous if its BCR is higher. The central tendency of the BCRs for Dam #1 is higher than that of Dam #2, per the simulation analysis that was done. As a result, Dam #1 is anticipated to produce, on average, more benefits per unit cost than Dam #2.

The BCRs of Dam #1 exhibit lower variability compared to those of Dam #2, indicating that the BCRs of Dam #1 are less widely dispersed than those of Dam #2. This implies that Dam #1 is more consistent in its performance and less prone to significant fluctuations in benefit-cost ratios. In terms of the distribution of BCRs, Dam #1 has a smaller skewness value than Dam #2, suggesting a more symmetrical or normal distribution of benefits and costs. This implies that the benefits and costs associated with Dam #1 are more evenly balanced than those of Dam #2.

The simulation analysis suggests that Dam #2 has a higher likelihood of generating extreme BCR values that surpass specific thresholds. This implies that Dam #2 may produce BCRs that are either exceptionally high or exceptionally low, which makes it a riskier option compared to Dam #1.

In summary, considering the factors mentioned above, it appears that Dam #1 is a safer and more dependable option than Dam #2. Although Dam #2 has the potential to produce very high BCRs, it also carries a greater risk of extreme values and variability, which could have negative consequences for the project's success. Conversely, Dam #1 is anticipated to deliver higher benefits per unit cost, and it is less likely to encounter significant fluctuations or extreme outcomes. Consequently, I advise selecting Dam #1 for the project.

**References:**

1. 1.3.6.6.6. Chi-Square Distribution. (n.d.). <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3666.htm>
2. Wittwer, J. (n.d.). Monte Carlo Simulation Basics. <https://www.vertex42.com/ExcelArticles/mc/MonteCarloSimulation.html>