Minerva Schools at KGI

Tuan Nguyen

CS112 Assignment

Question 1.

a. The mean duration of projects is 643.6 days, with the median of 592 days and interquartile range(IQR) of 384 days. When subsetting the dataset based on the circulation date, with earlier circulation date defined as the first 5 years and the later circulation date as the later 5 years, the result shows differences. For the earlier circulation date, the mean, median and IQR is 603.5071 days, 547.5 days, and 384.5 days respectively.

   Even though the mean is much larger than 18 months, the mean is not trustworthy because there might be some outliers affecting the mean result, such as a very long-duration project. Median, on the other hand, proves to be a better measurement. In this case, the median for both earlier and later circulation approximates 18 months. Therefore, we can infer that the statement is true.

b. The mean of planned project duration(P) and actual project duration(A) is around 643.6 and 1216.1 days respectively. The median and IQR of P is 592 and 384 days. The median and IQR for A is 1119 and 620.5 days respectively. There is a huge difference between revised date of completion versus scheduled time. The mean and IQR of P is actually half of A.

Question 2.

Rating value	Percentage(rounded to the nearest decimal)
0	2.7%
1	15.9%
2	68.3%
3	13.1%

The value of rating ranges from 0 to 3, with 0 represents the lowest success rate and 3 represents the highest success rate. In addition, all NA values (not available rating) have been removed. The most popular rating is 2, accounting for 68.3% of all ratings.

Question 3.

Rating value	Percentage(rounded to the nearest decimal)
0	2%
1	14%
2	70%
3	14%

The value of rating ranges from 0 to 3, with 0 represents the lowest success rate and 3 represents the highest success rate. The table shows the similar results with the first table. 2 is the most common rating, and it accounts for 70% of all ratings. Rating value 3 sees an increase of 1%, and there is an decrease in rating value 1, rounding to 14%.

Question 4.

	Top 25 % of RevisedAmount	Bottom 25% of RevisedAmount
Mean	1.93	1.91
Median	2	2
IQR	0	0

As p-value equals 0.33, which is not significant, it indicates that there is strong evidence against the null hypothesis. Therefore, we can not draw any causal conclusion about the effect of budget on rating.

Question 5.

a. Decision problem: optimize budget setting to maximize project's success. In other words, how to use/leverage budget/money to increase and help the project reach its success.

b. Levers: budget is the lever because we can change/ study the budget to make impact on the project's success.

c. RCT: Randomize to grant low amounts of loan and high amounts of loan to different projects.

d. Dependent variable: success of the project.

   Independent variable: the low/high amount of loan given to a project.

e. From RCTs, we can infer causal relationships between amount of money versus the success of the projects while other experiments such as observational and non-RCT experiments do not allow us to do so. The randomization also helps to mitigate any bias and provides highly credible evidence about the effect of an intervention.