GE 204 PROJECT 1 REPORT

1.

Normally, theoretically, when we roll a dice, the probability of getting numbers from 1 to 6 is 1/6 each. However, when we roll the dice 100 times in 2 steps, we see that the results are not exactly 1/6. Therefore, we can never expect the results to be the same as theoretical. Because terms and conditions differ. The more data sets we create, the more we reduce the error rate.

Based on my experiment results, the probabilities of rolling each number (1 through 6) do not exactly match the expected value of 1/6. The closest to the expected value is the possibility of rolling a 4, with a probability of 0.20 in the first 100 rolls and 0.16 in the second 100 rolls.

To calculate the percent error for each probability, I am using the formula:

- percent error = | (experimental probability - expected probability) / expected probability | * 100%

Using 1/6 as the expected value, the percent errors for each possibility in the first 100 rolls can be determined like these results:

- For 1: 20%
- For 2: 16.67%
- For 3: 20%
- For 4: 20%
- For 5: 16.67%
- For 6: 21.67%

In the second 100 rolls, the percent errors for each possibility can be determined like these results:

- For 1: 25%
- For 2: 25%
- For 3: 16.67%
- For 4: 6.67%
- For 5: 11.67%
- For 6: 16.67%

Comparing the percent errors obtained from the first 50 rolls and all 100 outcomes, it is possible that the percent error obtained from the first 50 rolls is larger, as it is based on a smaller sample size. However, this depends on the specific probabilities obtained from the first 50 rolls.

Overall, I celarly state that the experiment results do not perfectly match the expected probabilities, but the differences are not extreme. The largest percent error is for rolling a 1 or 2 in the second 100 rolls, but even this is within a reasonable range.

2.

In the second iteration of the histogram graph I made, I took the average of the numbers in every 5 iterations and tried to adapt it to 20 repetitions. The first histogram I made, I tried to show between which values the average numbers were.

The first histogram has the highest frequency around 3.5 and a somewhat symmetric shape. It also appears to be relatively spread out, with a range of values from around 2 to 6.

The second histogram does not appear to have a uniform distribution, which would show equal probability for each outcome. The histogram has the higher frequencies for outcomes 3, 4, and 6 and basically 1, and lower frequencies for outcomes 2, and 5. This distribution is consistent with the probabilities that were defined for each outcome.

The stem plot overlaid on the histogram shows the theoretical probabilities for each outcome, which do not match the observed frequencies perfectly but do follow the same general shape. Therefore, the observed data is consistent with the theoretical probabilities, but there may be some randomness or other factors that are causing the observed frequencies to deviate slightly from the theoretical probabilities.

As a result, the first histogram appears to be relatively normal while the second histogram is not uniform, but rather skewed towards certain outcomes. The second histogram is consistent with the theoretical probabilities that were defined for each outcome, and some deviation from these probabilities is expected due to randomness.