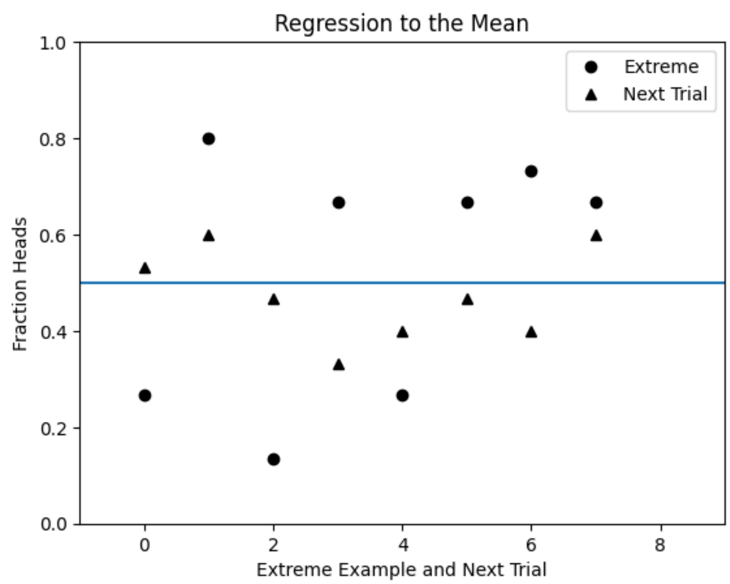
Henry Wandover

# Lab #3 Report

Probability, Distributions, and the Empirical Rule

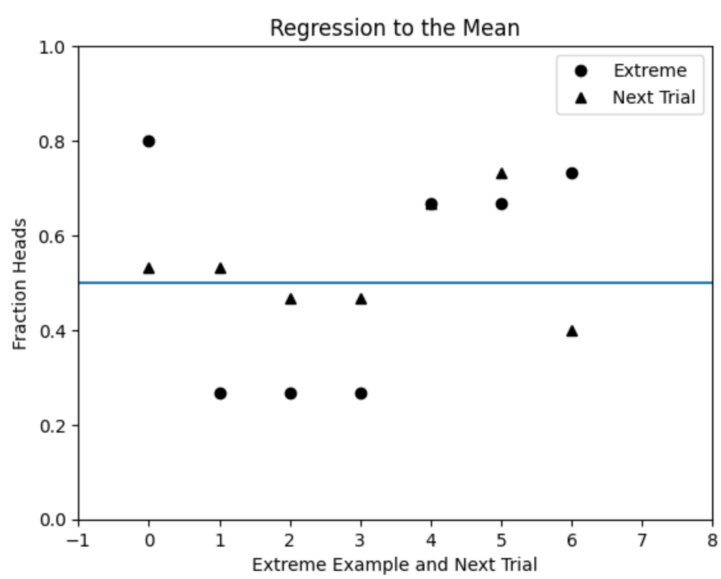
Part I

- I.I:



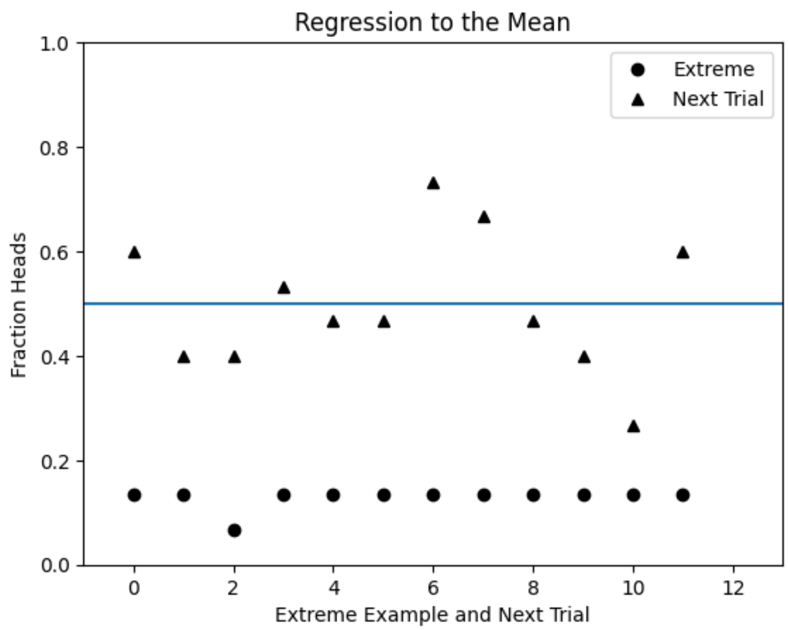

  - The blue line represents the mean of the given random data, looking at the graphed extremes and the following trial, each time the next trial appears closer to the mean. Except for the point at 3 on the x-axis which is reflected across the mean and is still the same distance from the mean.

- I.II

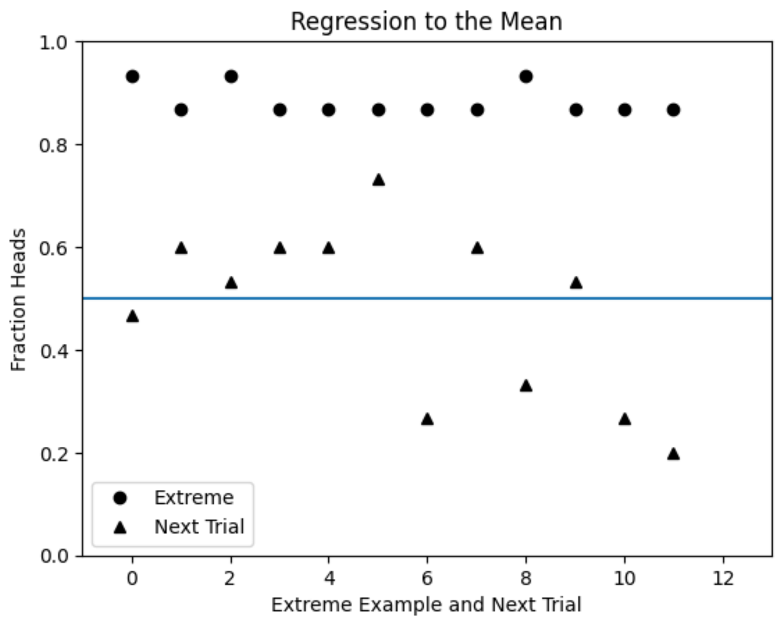



  - Here the point at 4 has an extreme and the next trial that is the same fraction of heads, and at 5 the extreme is closer to the mean. I believe the data is showing this way due to the gambler's fallacy, one would assume that after a certain amount of heads, a tail would land and then in the next trail would be closer to the mean. In a coin flip like this, there is no guaranteed outcome that can be mapped so to assume a regression to the mean is false.

- I.III




  - In this case I only accounted for extremes below 0.15, I had to run significantly more trials as it was unlikely that one would show with merely 40 trials. But when running more trials and after a couple of cycles it appears that with a lower extreme the frequency of regression to the mean was higher than normal.




- After looking at both, I can’t say that there was a significant enough difference in frequencies between an increase or decrease in the extreme value. Looking at the spread of patterns above, where the extreme accounted for was > 0.8, doesn’t lookm to dissimilar from the former. The lower extremes has a slightly tighter patter but increasing or decreasing show cases regression better than the normal model.

Part II

- II.I

  For mu = -3 and sigma = 3
  Fraction within 1 std = 0.6827
  Fraction within 1.96 std = 0.95
  Fraction within 2 std = 0.9545
  Fraction within 2.576 std = 0.99
  For mu = -7 and sigma = 3
  Fraction within 1 std = 0.6827
  Fraction within 1.96 std = 0.95
  Fraction within 2 std = 0.9545
  Fraction within 2.576 std = 0.99

- II.II
  - The two different trials don’t differ at all, since both examples are going to be normally distributed the mean = median = mode, and each time no matter if the mean and standard deviation differ the standard deviations are going to be the same. There is going to be complete symmetry around the center and 50% of values will be greater than the mean and 50% will be less everytime. In short terms, this is the empirical rule, 68%, 95%, and 99.7% of the values lie within one, two, and three standard deviations of the mean.
- II.III

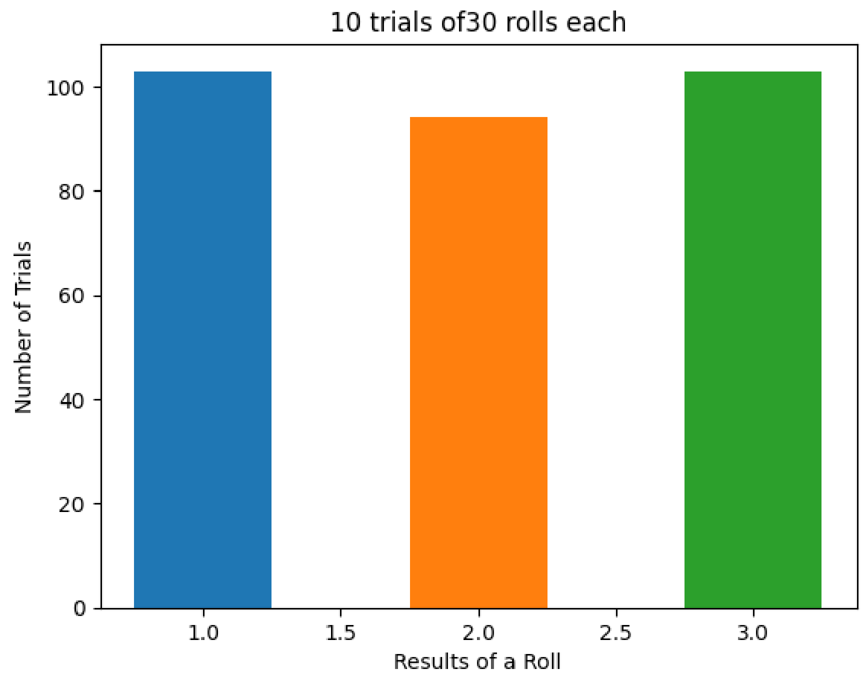
- The code calculates the given standard deviations from the given mean. To get there start with the gaussian function which takes a given point on the curve as an integer, the mean, and standard deviation. factor1 is set to the value of the height of the curve's peak which will be multiplied to factor2; that being the rest of the gaussian formula as seen, e to the power of negative integer minus the mean squared divided by two times the standard deviation squared.
- Then in the for loop the checkEmpirical applies each standard deviation into the gaussian formula using the quad function which finds a definite integral, with the limit between mean minus the number of standard deviations times the standard deviation and the mean plus the number of standard deviations times the standard deviation.
- That finds the area for each standard deviation and they are rounded up and it should be as appears based on empirical rule.

Part III

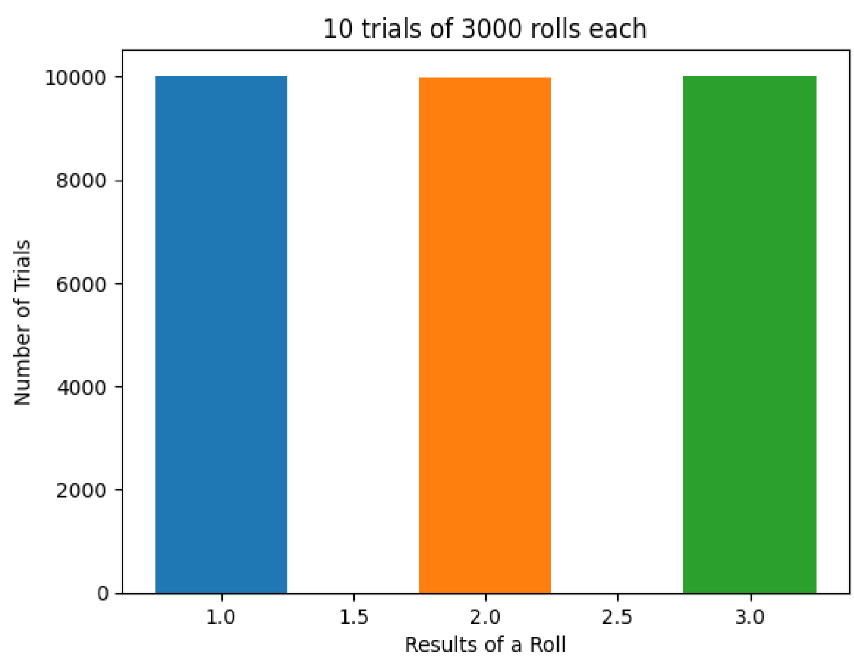
- III.I

```
def roll():
    """Assumes numRolls a positive int"""
    return random.choice((1, 2, 3))

def rollSim(numTrials, numRollsPerTrial):
    ones, twos, threes = 0, 0, 0
    for i in range(numTrials):
        for i in range(numRollsPerTrial):
            r = roll()
            if r == 1:
                ones += 1
            elif r == 2:
                twos += 1
            else:
                threes += 1
    return (ones, twos, threes)
```




- I would say with the trial of 30 rolls it's closest to a bimodal. Often it seems there is two more likely chances while one side is significantly less likely to appear. While not exactly bimodal it's the closest description.




- With the increase in trials to 3000 it's clear to see things even out and become much more of an uniform distribution as the number of trials has increased and it's safe to say the more trials the more uniform it will become.

- III.II

-

Part IV

- IV.I