

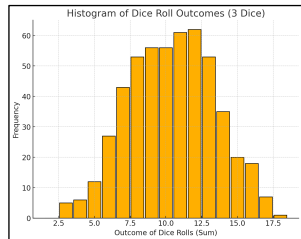
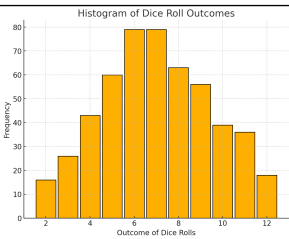
「サイコロを用いた統計データと正規分布の理論的整合性の研究」

林 理央(富山県立大門高等学校情報デザイン部 2 学年) 加藤 征馳(富山県立大門高等学校情報デザイン部 2 学年)
 坂上 宥詩(富山県立大門高等学校情報デザイン部 2 学年) 石崎 格杜(富山県立大門高等学校情報デザイン部 2 学年)

世間一般に公開されている統計情報の正しい読み方について基本的な知識やその数学的な背景を学び、その知識を実践的な場面で利用することができるようになるため、具体的な事例を選んで実際にアナログ的に実験をして調べてみた。母集団から抽出した標本から母集団の平均の範囲を推定することができる。その際に「95%信頼区間」での推定値を統計的に求めることができる。その95%とはどのような確率であるのか数学的に理解し、実際にサイコロを利用したアナログ手法を利用したシミュレーションで実際に確かめてみた。また、Python 及び表計算ソフトを利用したシミュレーションについて研究を進めた。

サイコロの出目の統計的検証

サイコロの目の出る確率は 1/6 である。本当にそうなるか実際にサイコロを 100 回 / 試行を 5 回繰り返して確率を調べた。サイコロを 2 個・3 個同時にふりそれぞれの合計値が二項分布に従うかシミュレーション実験を試みた。これまで学習した確率は数学的な裏付けがあると確認した。



正規分布する母集団の作成

様々な統計理論を検証するために、正規分布をする母集団サンプルを作成した。身近な題材とするためテストの点数分布を想定した要素数 100、平均 50、標準偏差 10 の母集団を作成した。

```
import numpy as np
import matplotlib.pyplot as plt

# 正規分布に従う乱数を生成
random_numbers = np.random.normal(loc=mean, scale=std_dev, size=size)
# 値を範囲[5, 95]内に収める
random_numbers = np.clip(random_numbers, 5, 95).astype(int)
# 生成された乱数を表示
print("生成された乱数:", random_numbers)

以下省略
```

標本平均から母平均を推定する

母集団から抽出した標本平均は正規分布に従うことを前提に、標本平均から母平均を 95% 信頼区間で推定することができる。また 95% の意味を検証することにした。

母平均の点推定と標本標準偏差の関係

正規分布する母集団から標本を抽出し、母平均を点推定し不偏統計値であることを確認した。また標本平均の標準偏差と母標準偏差と比較した。

	標本平均	標本平均の標準偏差
1	48.5	5.29
2	48.1	4.67
3	42.2	5.34
10	48.9	4.26

母集団の平均は 49.35 なのでほぼ標本平均と等しい。標本標準偏差は母標準偏差の $4^{1/2}$ 倍になっている。

標本数 4 の試行を 100 回を 10 回

母平均の区間推定と 95% 信頼区間

正規分布する母集団から標本を抽出し、母平均を 95% 信頼区間を推定し、母平均が推定範囲に含まれているかを検証した。検証は標本数 4 と 10 で実施し、100 および 500 回の検証をそれぞれ 10 回シミュレーションをした。(母分散未知)

母平均が推定範囲にある	範囲内にあった割合
959 回	95.9%

標本数 4 の試行を 100 回を 10 回

母平均が推定範囲にある	範囲内にあった割合
4784 回	95.7%

標本数 10 の試行を 100 回を 5 回

標本数 4、10 とも推定した範囲の中に実際に母平均が含まれている回数はそれぞれ標本数 4 の場合は 95.9%、標本数 10 の場合は、95.7% であった。95% 信頼区間の想定通りの確率で母平均が存在する範囲を推定していることを確認した。

母分散未知の場合の 95% 信頼区間推定範囲

$$\bar{x} - t_{\alpha/2}(n-1) \times \sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{x} + t_{\alpha/2}(n-1) \times \sqrt{\frac{s^2}{n}}$$

x は標本平均 s は標本分散