Lab4

April 26, 2023

```python
[1]: import numpy as np
     import pandas as pd
```

0.1 1. Test Data Under H0

1.1. What is your H0–the claim about the world you want to test? Explain!   The null hypothesis I want to test is more than 50% of the people prefer red. Since the expected value of red voters is less than 0.5, there is 95% confidence that we can reject the null hypothesis.

1.2. Choose the number of polls (repetitions R) you conduct. 1000 is a good choice.

```python
[2]: R = 1000;
     p = 0.5;
     N = 1000;
```

1.3. Now conduct R polls of N respondents. Each respondent should be represented as 0 or 1, with 1 occuring with the probability 0.5 (your H0).

```python
[3]: np.random.seed(1)
     data_mean = []
     for i in range (R):
         poll = np.random.binomial(1, p, size=N)
         data_mean.append(np.mean(poll))
```

1.4. Find the 95% CI of your means by computing 2.5-th and 97.5-th percentiles.

```python
[4]: p_2_5, p_97_5 = np.percentile(data_mean, (2.5, 97.5))
     print("2.5th Percentile: " + str(p_2_5))
     print("97.5th Percentile: " + str(p_97_5))
```

```
2.5th Percentile: 0.469
97.5th Percentile: 0.533
```

1.5. Does your actual poll result (0.53) fall into the CI?   Yes, 0.53 falls into the confidence interval,

1.6.Based on the CI you computed answer the question: can you confidently (at 5% significance level) say that Greens are leading in the polls?   Yes, hence we can reject the null hypothesis and say that I am 95% confident that Greens are leading the poll.

0.2 2. Test H0 under data

2.1. Now conduct R polls of N respondents. Each respondent should be represented as 0 or 1, with 1 occuring with the probability 0.53 (this is the probability in data). For each poll, compute the average value (the proportion of respondents who prefer G). Now you should have R averages.

```python
[5]:  p = 0.53
      np.random.seed(1)
      data_mean = []
      for i in range (R):
          poll = np.random.binomial(1, p, size=N)
          data_mean.append(np.mean(poll))
```

2.2. Find the 95% CI of your means by computing 2.5-th and 97.5-th percentiles.

```python
[6]:  p_2_5, p_97_5 = np.percentile(data_mean, (2.5, 97.5))
      print("2.5th Percentile: " + str(p_2_5))
      print("97.5th Percentile: " + str(p_97_5))
```

```
2.5th Percentile: 0.497
97.5th Percentile: 0.561
```

2.3. Is 0.5—your H0 inside of the CI?   Yes, 0.5 is within the confidence interval

2.4. Based on the CI you computed answer the question: can you confidently (at 5% significance level) say that Greens are leading in the polls?   Yes I can confidently say at 5% significance level that Greens are leading the polls.