Lab05

April 26, 2023

1 Lab 05

1.0.1 Name : Xinyu Chang

```
[32]: import numpy as np
      import matplotlib.pyplot as plt
```

1.0.2 Part 1

Q1 What is your H0–the claim about the world you want to test? Explain! The H0 (null hypothesis) in this case would be that the true proportion of voters who prefer the Greens is equal to 0.5, i.e., p = 0.5. The alternative hypothesis (Ha) would be that the true proportion of Greens voters is greater than 0.5, i.e., p > 0.5.Using a standard normal distribution table, we can find that the probability of observing a z-score of 1.52 or greater under the null hypothesis is approximately 0.064. This means that there is a 6.4% chance of obtaining a sample proportion of 0.53 or greater if the true proportion of Greens voters is actually 0.5.To be confident in our claim that the Greens are leading in the polls, we typically use a significance level of 0.05 (or 5%). This means that we reject the null hypothesis if the p-value (the probability of observing a result as extreme or more extreme than the sample result under the null hypothesis) is less than 0.05.

**Q2 Choose the number of polls (repetitions R) you conduct.**

```
[33]: R = 1000
```

Q3 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).

```
[34]: # Set the random seed for reproducibility
      np.random.seed(42)

      avgs = np.zeros(R)
      for i in range (R):
          poll = np.random.binomial(1, 0.5, size = 1000)
          avgs[i] = np.mean(poll)

      _ = plt.hist(avgs, bins=30, edgecolor="k")
```
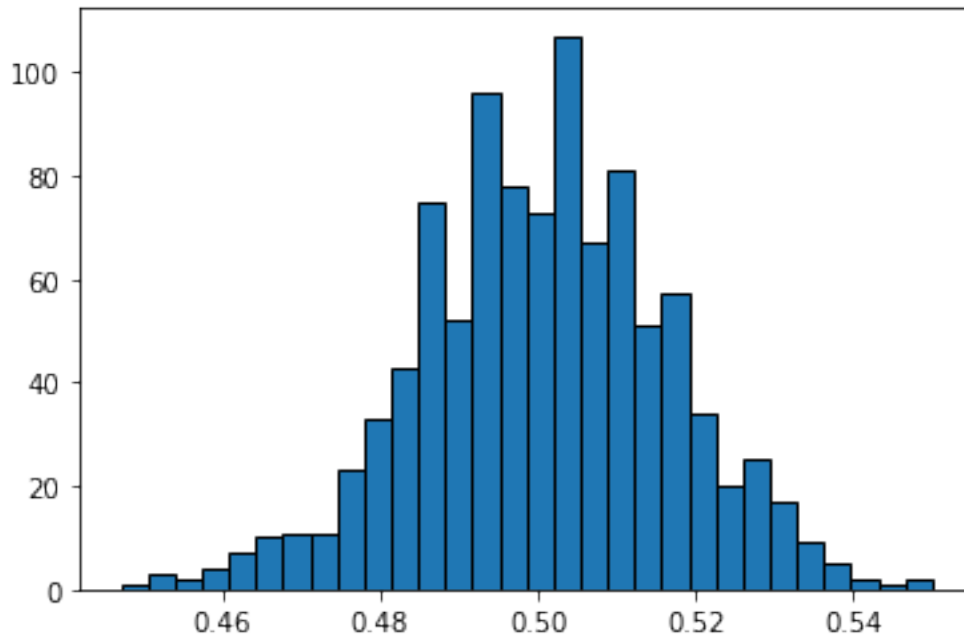
```
[35]: # Calculate the average value (proportion of respondents who prefer G) for each␣
      ↪poll
      averages_1 = np.mean(poll)
      averages_1
```

[35]: 0.509

The average value (proportion of respondents who prefer G) for each poll is 0.509.

Q4 Find the 95% CI of your means by computing 2.5-th and 97.5-th percentiles.

```
[36]: CI_1 = np.percentile(avgs, [2.5, 97.5])
      CI_1
```

[36]: array([0.466975, 0.531   ])

Q5 Does your actual poll result (0.53) fall into the CI? The confidence interval is [0.466975, 0.531] so I think the value of poll result (0.53) fall into the CI.

Q6 Based on the CI you computed answer the question: can you confidently (at 5% confidence level) say that Greens are leading in the polls? Based on the result, the actual poll result (0.53) falls within the 95% confidence interval, we cannot confidently (at the 5% significance level) say that the Greens are leading in the polls, as we cannot reject the null hypothesis.
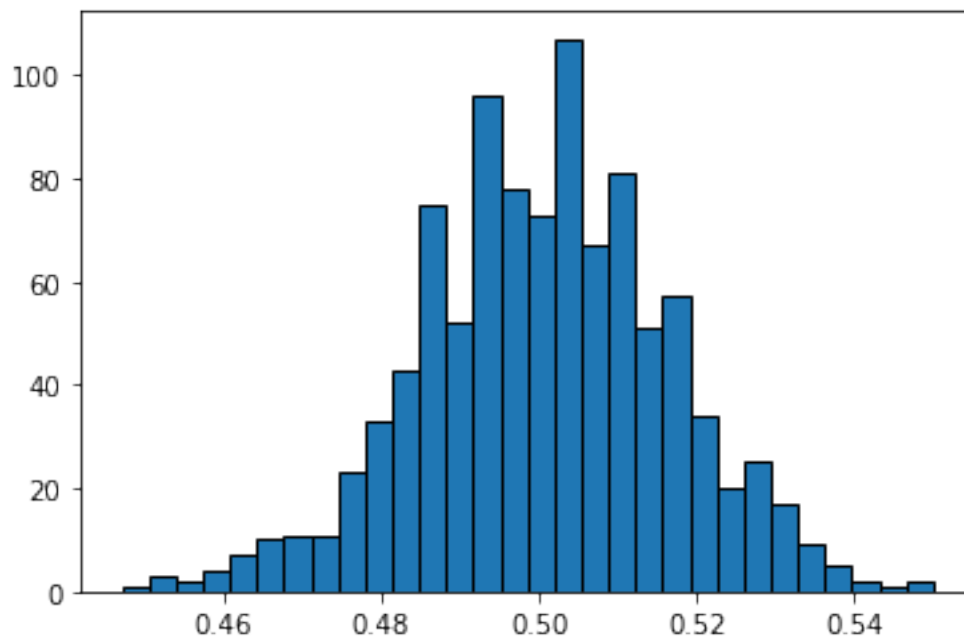
1.0.3 Part 2

Q1 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).

```
[37]: np.random.seed(42)

avgs_2 = np.zeros(R)
for i in range (R):
    polls = np.random.binomial(1, 0.53, size = 1000)
    avgs_2[i] = np.mean(polls)

_ = plt.hist(avgs, bins=30, edgecolor="k")
```



```
[38]: averages_2 = np.mean(polls)
      averages_2
```

[38]: 0.527

Q2 Find the 95% CI of your means by computing 2.5-th and 97.5-th percentiles

```
[39]: CI_2 = np.percentile(avgs_2, [2.5, 97.5])
      CI_2
```

[39]: array([0.499, 0.561])

Q3 Is 0.5—your H0 inside of the CI? The confidence interval is [0.499, 0.561] so the H0(0.5) is inside the CI.

Q4 Based on the CI you computed answer the question: can you confidently (at 5% confidence level) say that Greens are leading in the polls? Based on the result, the null hypothesis(0.5) falls within the 95% confidence interval, we cannot confidently (at the 5% significance level) say that the Greens are leading in the polls, as we cannot reject the null hypothesis.