

Data 8
Worksheet 3
Conceptual Office Hours

Name: _____

This worksheet can serve as a general overview of materials you have learned this week and give you an opportunity to practice solving them in an exam-like format. NOTE: This is not necessarily representative of what will be tested on the actual exams. For any issues or questions, contact Robert Sweeney Blanco at robertsweeneyblanco@berkeley.edu.

For all problems, you may assume you have imported datascience and numpy as np.

You're at The Wizarding World of Harry Potter at Universal Studios Hollywood where a wizard sells you an enchanted coin that has a 40% chance of landing heads. When you get home, you flip the coin 100 times and get 50 heads. You begin to think the wizard is a fraud. Design a hypothesis test to determine whether it is believable that the coin is enchanted. (Some of these questions have multiple possible answers)

1. What is your Null Hypothesis?

2. What is your Alternative Hypothesis?

3. What is your test statistic?

4. Fill in the code to run the simulation. Note: The answers to these depend on your answers above!

```
proportions = <1>
collection = <2>
for i in np.arange(1000):
    sample = 100 * sample_proportions(<3>, proportions).item(<4>)
    collection = <5>
```

4.1) _____

4.2) _____

4.3) _____

4.4) _____

4.5) _____

5. Assuming you have access to *collection* from your code above. Write code to compute the p-value.

6. Pretend the line above evaluated to .024. Using the standard p-value cutoff of 0.05, is this result statistically significant or not?

7. What is your conclusion with regards to whether the coin is enchanted or not?

France has been trying to improve its consumption of clean energy. As a data scientist, you want to know if France's efforts have paid off, or if their energy consumption is like the rest of Europe's and variation is due to random chance. You have access to the table below, called **energy**. Additionally, you know that France consumed approximately 2,826,000 kilowatt-hours of energy.

Energy Source	Europe Average	France
Oil	0.34	0.28
Gas	0.20	0.16
Coal	0.16	0.04
Renewables	0.13	0.12
Nuclear	0.13	0.40

8. What would be your Null Hypothesis?

9. What would be your Alternative Hypothesis?

10. What statistic would you use to test the hypothesis? Why?

11. Fill in the function to compute the test statistic. Call it on **energy**. Use as many arguments as you need, and fill in each line corresponding to the numbers you see in the code.

```
def test_stat(dist1, dist2):  
    return <1>(np.abs(<2>)) / <3>
```

```
observed_value = test_stat(<4>, <5>)
```

11.1) _____

11.2) _____

11.3) _____

11.4) _____

11.5) _____

12. Write code to help you simulate one value of the test statistic.

```
total_energy = 2826000  
random_distribution = <1>(<2>, <3>)  
simulated_value = <4>(<5>, <6>)
```

12.1) _____

12.2) _____

12.3) _____

12.4) _____

12.5). _____

12.6) _____