Distribution Stations

In nature, events follow different probability patterns. For example, what is the probability of getting heads when tossing a coin? What is the probability that an earthquake will occur within the next 10 years? If we vaccinate 200 people, what is the probability that at least 100 of them develop immunity? Since these events follow different patterns, statisticians have defined various *probability distributions* to model and predict occurrences. You may already be familiar with the *normal distribution*, but in statistics, there are several other important distributions. It might be helpful to revisit probability and distributions from your first-semester coursework.

In this challenge, you have to estimate the distributions of four different populations by collecting samples from our Station's.

Stations

Across the JK building, we have placed four Station's. Every Station represents a population, and it prompts a new piece of data from that population every 10 seconds.

The locations of these Station's are provided to you during today's plenary introduction. Here is a reminder:

- Station 1 - JK 1.11
- Station 2 - JK 1.26
- Station 3 - JK 3.13
- Station 4 - somewhere in CSSci Wing

Your Task

Using whatever digital or manual tool you have:

1. Plot the sample data you collected together with the estimated **probability distribution functions (pdf's)** (or pmf's) of all four Station's.

2. Provide the type of the estimated distribution together with the estimated *parameter* values for each Station.

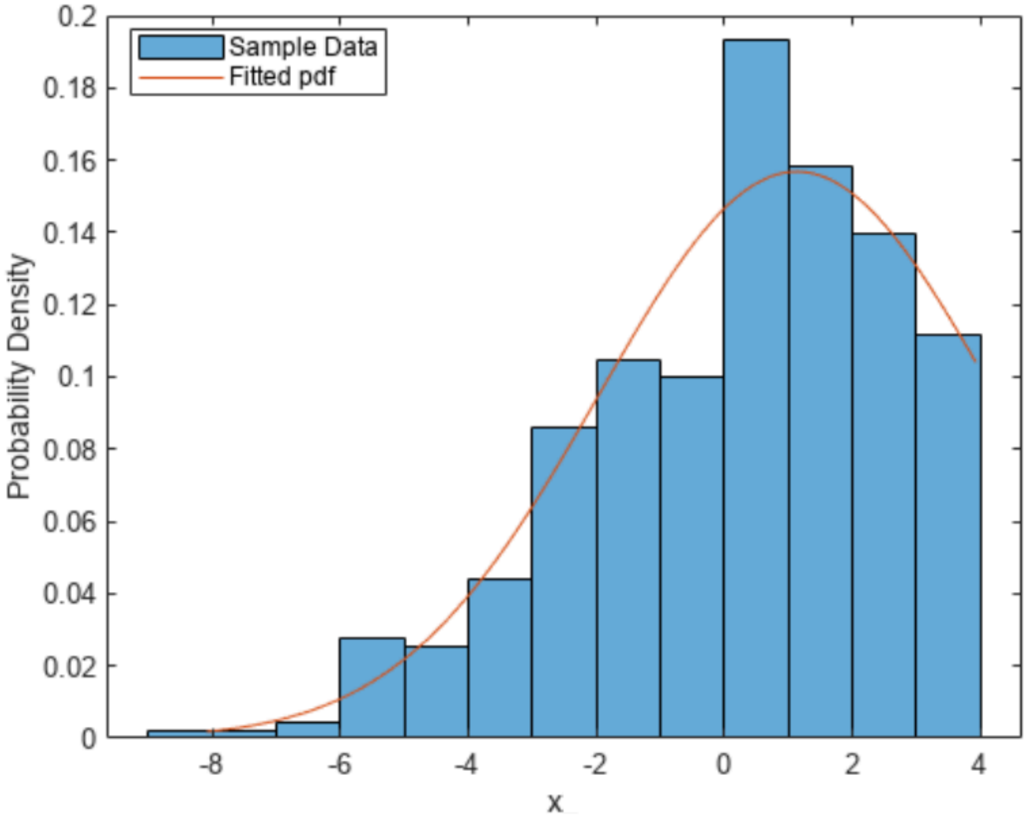
Hints

- Larger samples you collect, more *accurately* you'll be able to estimate their distribution. Why?

- You learned many types of data visualizations last semester. Which type(s) enable you to quickly visualize *the distribution of a data sample*? Think about this: What would you want to see in your y-axis and x-axis?

- Precision level of your x-axis is critical in detecting the distribution type. Check the example visual below. How could you play with the precision of x-axis in this type of graph?

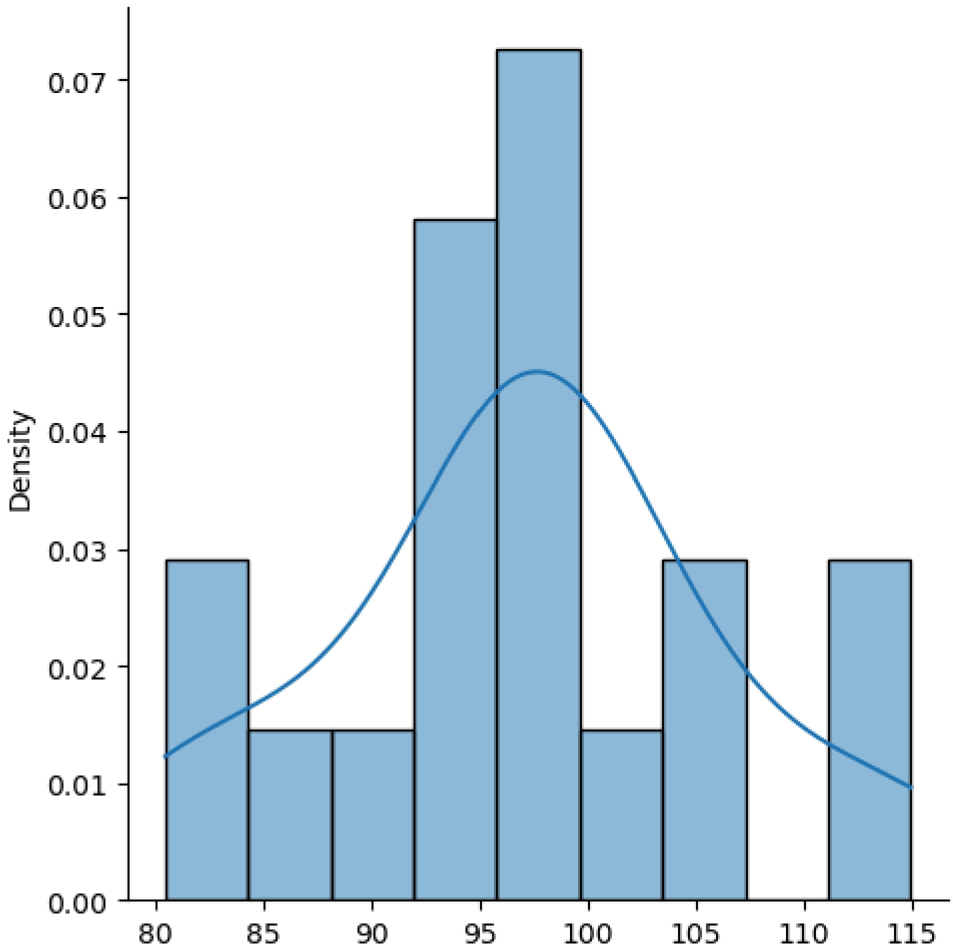
- Here is an example figure plotting both the sample data and estimated distribution. Aim to generate a similar plot comparing the sample you collected and distribution you estimate for each Station to be able to judge *the goodness of your estimate*.



In [1]:
```
import seaborn as sns
```

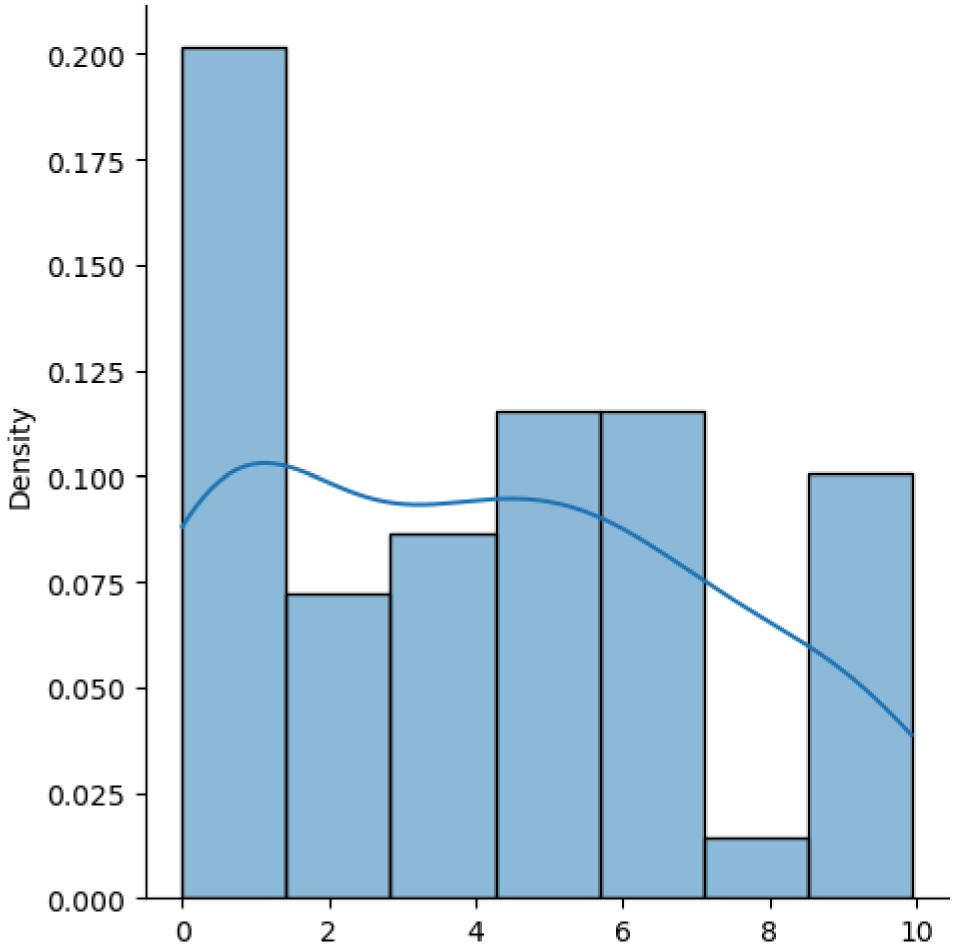
In [28]:
```
station_1 = [99.27942, 104.64406, 86.08853, 80.48506, 99.12515, 114.93615, 97.25
sns.displot(station_1, stat = 'density', kde = True)
```

Out[28]: <seaborn.axisgrid.FacetGrid at 0x20d6eab9f70>



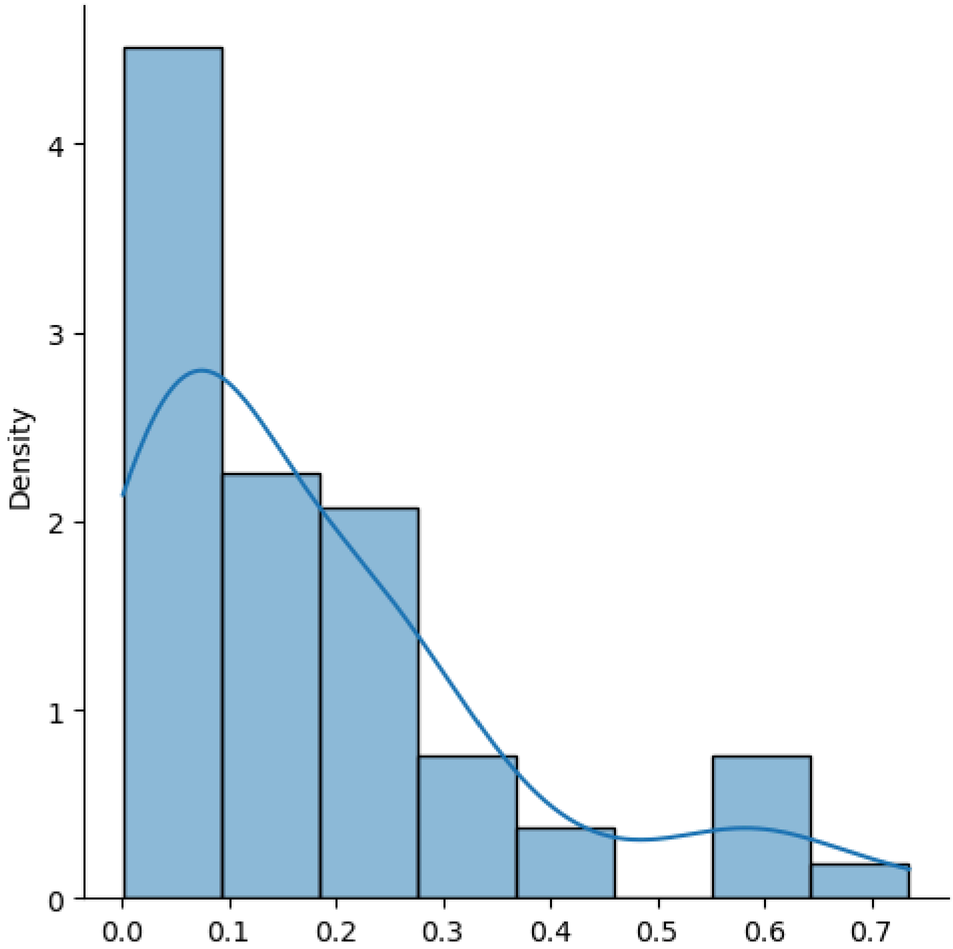
In [29]:
```
station_2 = [8.63407, 0.74968, 1.67941, 3.93575, 0.48684, 6.05979, 6.57805, 0.02
sns.displot(station_2, stat = 'density', kde = True)
```

Out[29]: `<seaborn.axisgrid.FacetGrid at 0x20d6d88de50>`



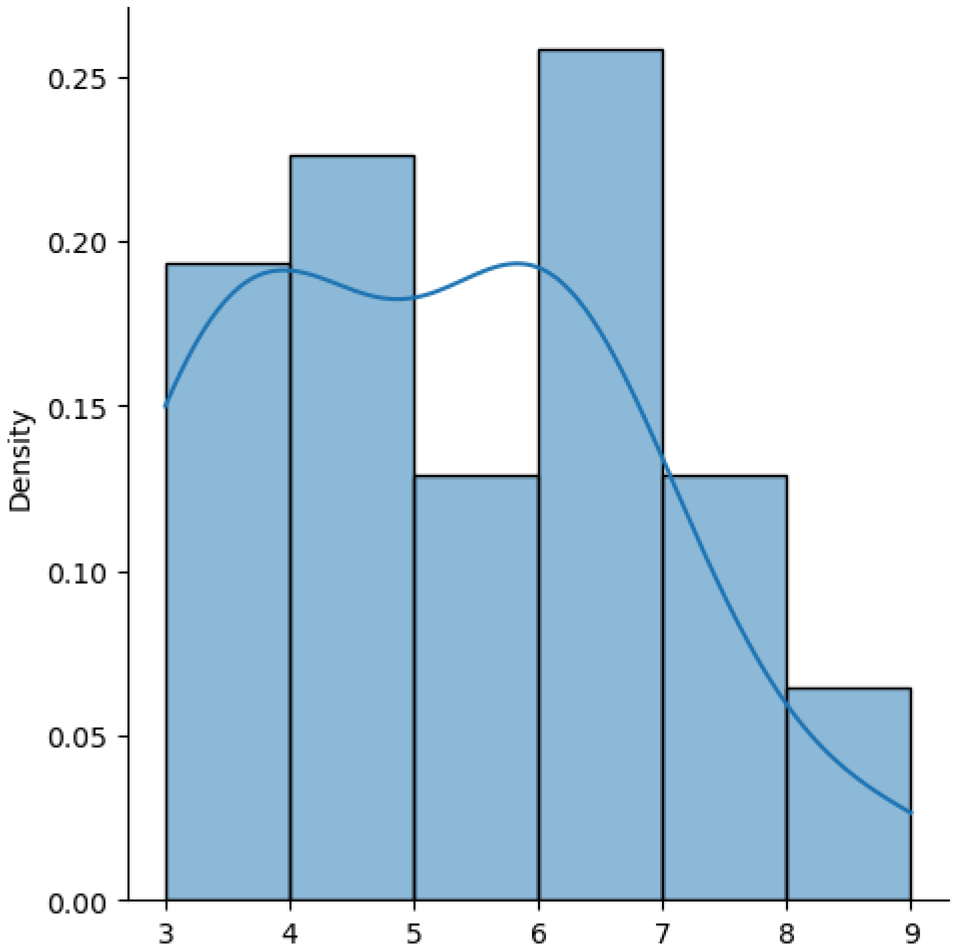
In [30]:
```
station_3 = [.17648, .55467, .01952, .05784, .07216, .08006, .14580, .01113, .13
sns.displot(station_3, stat = 'density', kde = True)
```

Out[30]:
```
<seaborn.axisgrid.FacetGrid at 0x20d6e9bfd70>
```



In [31]:
```
station_4 = [5, 9, 3, 6, 4, 6, 3, 4, 7, 7, 6, 5, 3, 4, 7, 5, 4, 3, 3, 6, 6, 8, 4
sns.displot(station_4, stat = 'density', kde = True)
```

Out[31]: <seaborn.axisgrid.FacetGrid at 0x20d72779ac0>



In [ ]: