Dakota Wilson

DSC-510 T1DQ2

1/29/2024

Loom video link: https://www.loom.com/share/3c4fad35b01f4092b6ece8c68f02b1f5

There are various ways that data can be misleading or misinterpreted. One example of this is seen with sample bias. If a sample is biased and does not accurately represent the desired population, results from the study may not actually apply to the population or be misinterpreted. One simple way Python can be used as a tool to help determine this is to see if the sample statistically matches the demographic profile of our population. It is important to understand that just because the demographic profile matches our expectation it does not confirm our sample is an accurate representation of our polulation. If our sample contains ethnic or gender data we can compare that to what we should expect in the area. We can also look at gender data to identify trends in gender among specific diagnoses. For example, lets take the gender distribution of some sample data and compare it to what we would expect to see with the gender distribution of children diagnosed with Autism. According to Posserud et al. (2021), the male to female ratio of children with autism was 3.67 to 1. We will see how our sample data compares to that.

In [19]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

I will generate a sample of 1000 patient genders of diagnosed autism children and compare the demographics of that sample to see if it fits the expected demographics of our population.

In [27]:
```
# generate sample gender data
patient_id = range(1, 1001) #assign patient ID 1-1000
sample_gender = np.random.choice(['Male', 'Female'], size = 1000) # randomize gender

#DataFrame of sample data with patient ID and gender
sample_data = pd.DataFrame({'ID': patient_id, 'Gender': sample_gender})

print(sample_data)
```

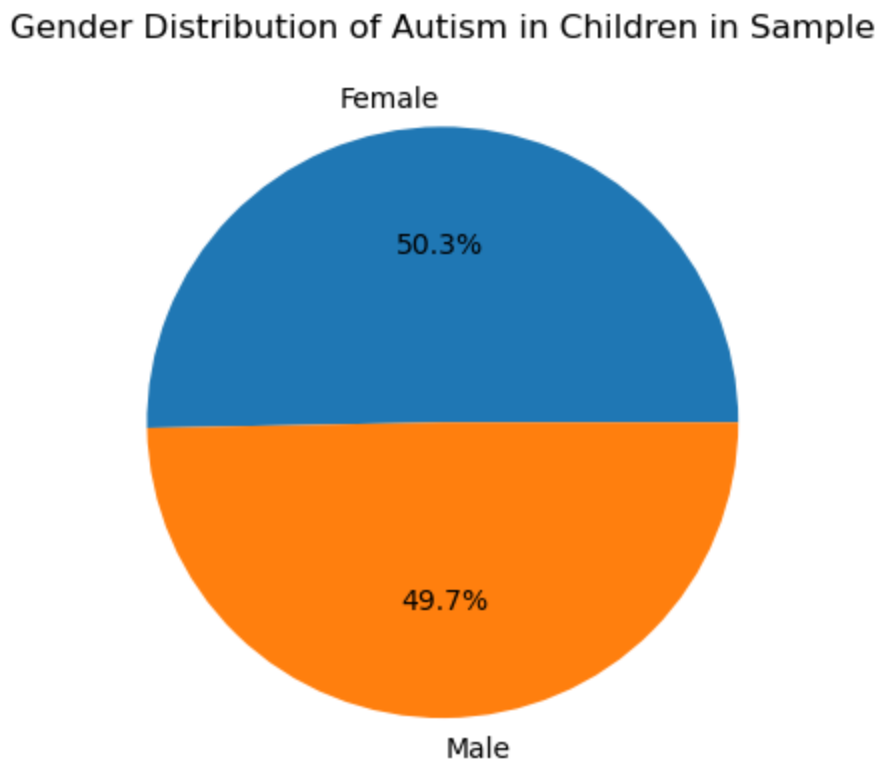
```
        ID  Gender
0        1    Male
1        2  Female
2        3  Female
3        4  Female
4        5  Female
..     ...     ...
995    996    Male
996    997  Female
997    998    Male
998    999  Female
999   1000  Female

[1000 rows x 2 columns]
```

In [28]:
```
# count occurences of each gender in sample
sample_gender_count = sample_data.groupby('Gender').size()

print(sample_gender_count)
```

```
Gender
Female    503
Male      497
dtype: int64
```

In [29]:
```
# percentage of each gender in sample
sample_gender_percent = (sample_gender_count.values/sample_gender_count.sum())*100
```

In [30]:
```
# Sample Gender Distribution pie chart
plt.pie(sample_gender_percent, labels=sample_gender_count.index, autopct='%1.1f%%')
plt.title('Gender Distribution of Autism in Children in Sample')

plt.show()
```



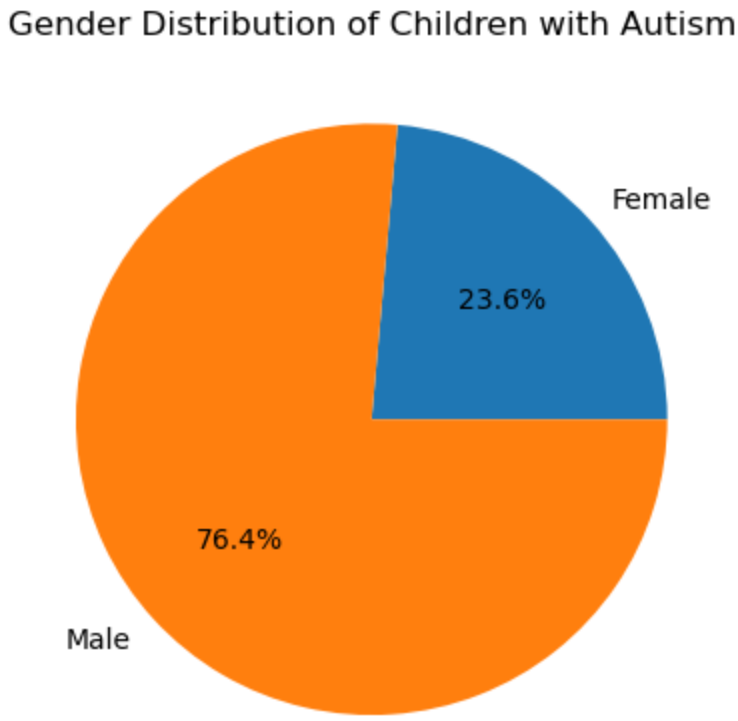
This pie chart shows an even distribution of genders for the sample data. For reference, I will create a pie chart of a sample generated with the expected gender distriibution.

In [31]:
```
# calculate gender probabilitity values
probability_male = 3.67/4.67
probability_female = 1/4.67

# generate autism gender data
autism_gender = np.random.choice(['Male', 'Female'], size = 1000, p=[probability_male,

# DataFrame of autism data with patient ID and gender
autism_data = pd.DataFrame({'ID': patient_id, 'Gender': autism_gender})
```

In [32]:
```
# count tally of each gender in autism data
autism_gender_count = autism_data.groupby('Gender').size()

print(autism_gender_count)
```

```
Gender
Female    236
Male      764
dtype: int64
```

In [33]:
```
# percentage of each gender
autism_gender_percentage = (autism_gender_count.values/autism_gender_count.sum())*100

# pie chart of referenced population data
plt.pie(autism_gender_percentage, labels=autism_gender_count.index, autopct='%1.1f%%')
plt.title('Gender Distribution of Children with Autism')

plt.show()
```



This pie chart represents what would be expected for a sample that represents the population of children with autism.

References:

Posserud, M.-B., Skretting Solberg, B., Engeland, A., Haavik, J., & Klungsøyr, K. (2021). Male to female ratios in autism spectrum disorders by age, intellectual disability and attention-deficit/hyperactivity disorder. Acta Psychiatrica Scandinavica, 144(6), 635–646. https://doi.org/10.1111/acps.13368