Understanding Central Tendency

In [3]:
```
Central tendency is the property of data to be distributed about a characteristic v
most important measures of central tendency are the mean and median.
```

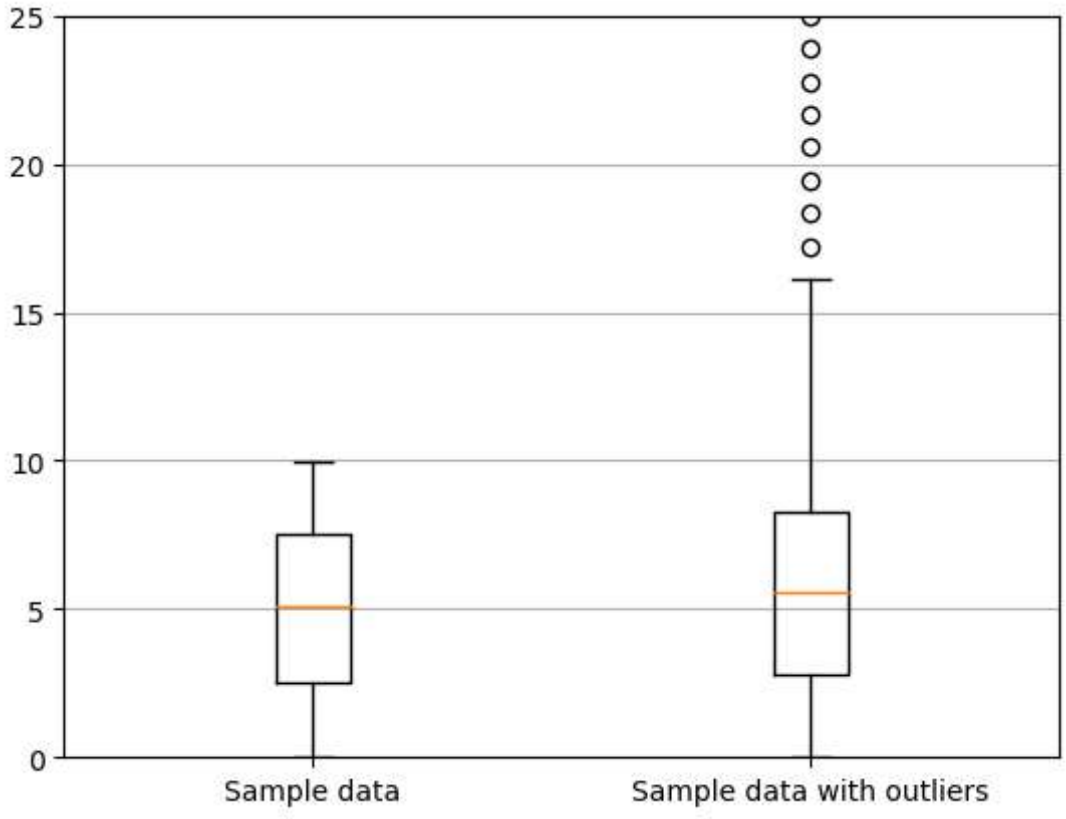
Calculating Mean and Median for a Dataset

In [14]:
```
import numpy as np
import matplotlib.pyplot as plt
```

In [8]:
```
np.random.seed(1)
data1=np.random.uniform(0,10,1000)
data2=np.append(data1,np.linspace(150,200,100))
data2=np.append(data2,np.linspace(15,25,10))
data=list([data1,data2])
fig,ax=plt.subplots()
#build a box plot
ax.boxplot(data)
ax.set_ylim(0,25)
xticklabels=['Sample data','Sample data with outliers']
ax.set_xticklabels(xticklabels)

#add horizontal grid lines
ax.yaxis.grid(True)

#Show the plot
plt.show()
```



Mean and Median of sample data with out outliers

In [10]:
```
np.mean(data1)
```

Out[10]:
```
5.006045994559051
```

In [11]:
```
np.median(data1)
```

Out[11]:
```
5.075008116147119
```

Mean and Median of sample data with outliers

In [12]:
```
np.mean(data2)
```

Out[12]:
```
20.455897292395537
```

In [13]:
```
np.median(data2)
```

Out[13]:
```
5.565300519330409
```

We observe that the presence of outliers in the second dataset led to an increase in the mean value from 5.56 to 20.45, while the change in median value from 5.07 to 5.0 was very small compared to the change in the man value.

This shows that the median value is a robust measure of central tendency as it is less susceptible o the presence of outlers in the dataset

Summary

in summary, we have reviewed the two most important metrics for calculating central tendency. The mean value is easy to compute, but is highly susceptible to the presence of outliers in the dataset. The median is a robus measure of central tendency, and is less susceptible to the presence of outliers.

In [ ]:
```
```