

**DEPARTMENT OF INFORMATION TECHNOLOGY**

# COURSE CODE: DJS22ITL5013

**COURSE NAME:** Statistical Analysis Lab **CLASS:** T.Y. BTech

**NAME:** Anish Sharma

# EXPERIMENT NO.02

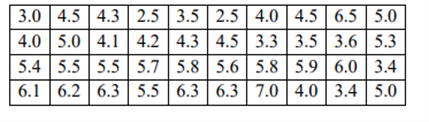
**CO 1:** Interpret the data using Descriptive Statistics.

**AIM / OBJECTIVE:** To explore descriptive statistics- Measures of Central Tendency

# DESCRIPTION OF EXPERIMENT:

**Perform the following**

1. Weight of babies (kg) below 6 months taken from a hospital record is given below. Calculate Mean, Median, Mode.



import numpy as np

from scipy import stats

# Data

weights = [3.0, 4.5, 4.3, 2.5, 3.5, 2.5, 4.0, 4.5, 6.5, 5.0, 4.0, 5.0, 4.1, 4.2, 4.3, 4.5,

           3.3, 3.5, 3.6, 5.3, 5.4, 5.5, 5.5, 5.7, 5.8, 5.6, 5.8, 5.9, 6.0, 3.4, 6.1, 6.2,

           6.3, 5.5, 6.3, 6.3, 7.0, 4.0, 3.4, 5.0]

# Mean

mean\_weight = np.mean(weights)

# Median

median\_weight = np.median(weights)

# Mode

mode\_result = stats.mode(weights, keepdims=False)

mode\_weight = mode\_result.mode  # Access mode value directly

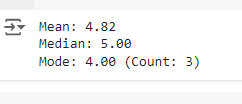
mode\_count = mode\_result.count  # Access count value directly

print(f"Mean: {mean\_weight:.2f}")

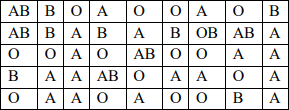
print(f"Median: {median\_weight:.2f}")

print(f"Mode: {mode\_weight:.2f} (Count: {mode\_count})")

What are your observations about the data, based on the central tendency values.



1. Draw a Bar diagram, Pie chart for the following data on the blood group of 45 students in a class.



import matplotlib.pyplot as plt

# Data

blood\_groups = ['AB', 'B', 'O', 'A']

counts = [4, 13, 12, 14]

# Bar Diagram

plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)

plt.bar(blood\_groups, counts, color=['red', 'blue', 'green', 'purple'])

plt.xlabel('Blood Group')

plt.ylabel('Number of Students')

plt.title('Bar Diagram of Blood Groups')

# Pie Chart

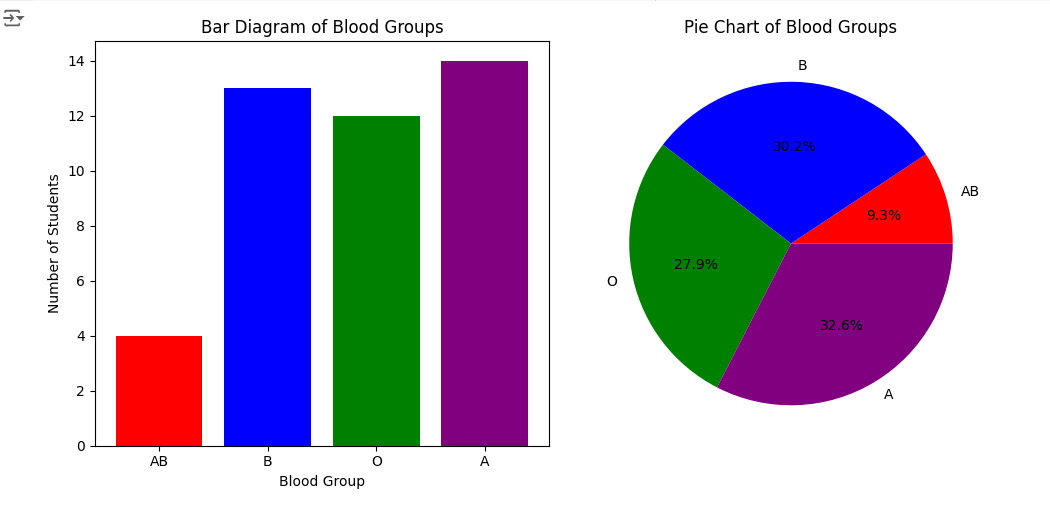
plt.subplot(1, 2, 2)

plt.pie(counts, labels=blood\_groups, autopct='%1.1f%%', colors=['red', 'blue', 'green', 'purple'])

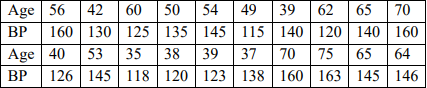
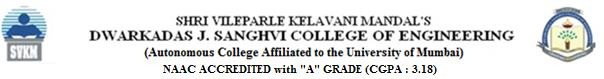
plt.title('Pie Chart of Blood Groups')

plt.tight\_layout()

plt.show()



1. Draw a line graph for age(years) versus systolic blood pressure (mm Hg) for the following data given below



import matplotlib.pyplot as plt

# Data

ages = [56, 42, 60, 50, 54, 49, 39, 62, 65, 70, 40, 53, 35, 38, 39, 37, 70, 75, 65, 64]

bps = [160, 130, 125, 135, 145, 115, 140, 120, 140, 160, 126, 145, 118, 120, 123, 138,

       160, 163, 145, 146]

# Plot

plt.figure(figsize=(10, 6))

plt.plot(ages, bps, marker='o', linestyle='-', color='b')

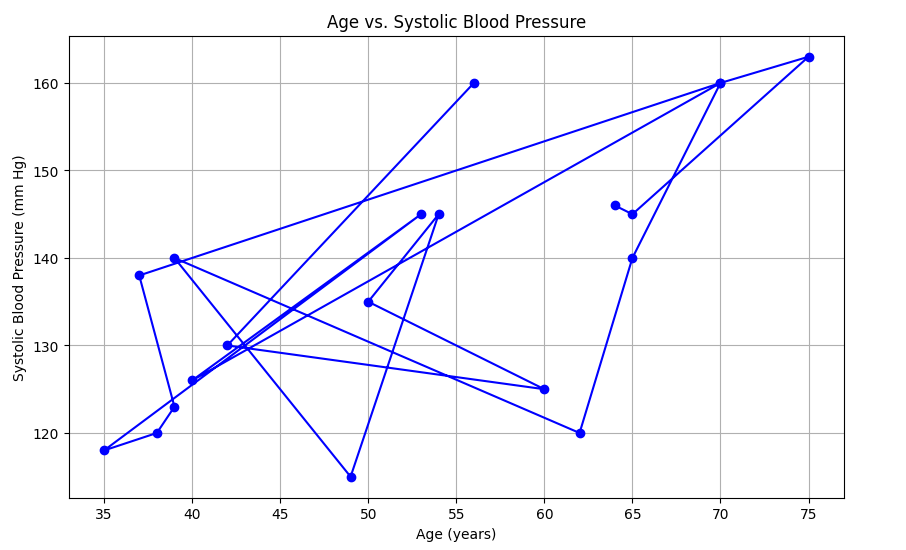
plt.xlabel('Age (years)')

plt.ylabel('Systolic Blood Pressure (mm Hg)')

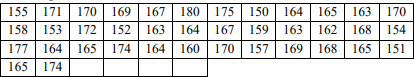
plt.title('Age vs. Systolic Blood Pressure')

plt.grid(True)

plt.show()



1. Draw a histogram with frequency curve for the following data on height in cm



import matplotlib.pyplot as plt

import seaborn as sns

# Data

heights = [155, 171, 170, 169, 158, 153, 172, 152, 177, 164, 165, 174, 165, 174, 167, 180,

           175, 150, 164, 165, 163, 170, 163, 164, 164, 167, 159, 163, 162, 168, 154, 160,

           170, 157, 169, 168, 165, 151]

# Plot

plt.figure(figsize=(12, 6))

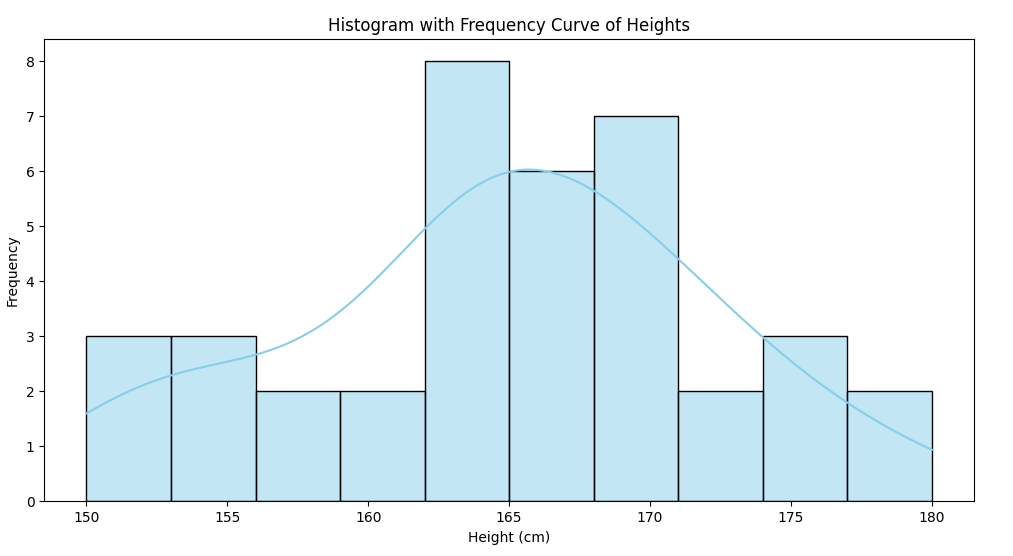
sns.histplot(heights, bins=10, kde=True, color='skyblue')

plt.xlabel('Height (cm)')

plt.ylabel('Frequency')

plt.title('Histogram with Frequency Curve of Heights')

plt.show()



1. Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies (f) that were obtained for different values of x, the number of heads, are shown in the following table. Calculate median, quartiles, and 27th percentile.



import numpy as np

# Data

heads = [0, 1, 2, 3, 4, 5, 6, 7, 8]

frequencies = [1, 9, 26, 59, 72, 52, 29, 7, 1]

# Create dataset from frequencies

data = np.repeat(heads, frequencies)

# Median

median = np.median(data)

# Quartiles

Q1 = np.percentile(data, 25)

Q3 = np.percentile(data, 75)

# 27th Percentile

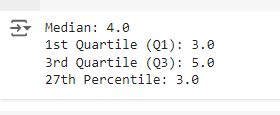
percentile\_27 = np.percentile(data, 27)

print(f"Median: {median}")

print(f"1st Quartile (Q1): {Q1}")

print(f"3rd Quartile (Q3): {Q3}")

print(f"27th Percentile: {percentile\_27}")



1. Find the mean, median and mode for the following collection of responses to the question: "How many parking tickets have you received this semester?"

1, 1, 0,1, 2, 2, 0, 0, 0, 3, 3,0, 3, 3, 0,2, 2, 2, 1, 1,4, 1, 1,0,3, 0, 0, 0, 1, 1, 2, 2, 2, 2,1, 1, 1, 1,

4, 4, 4,1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,1, 1, 1, 1, 1, 3,3,0, 3, 3, 1, 1, 1, 1,0, 0, 1, 1, 1, 1, 3, 3,

3, 2, 3, 3, 1, 1, 1,2, 2, 2,4, 5, 5, 4, 4, 1, 1, 1, 4,1, 1, 1,3, 3, 5,3, 3, 3, 2,3, 3, 0, 0, 0, 0, 3, 3, 3,

3, 3, 3, 0, 2, 2, 2, 2, 1, 1, 1,3, 1, 0, 0, 0,1, 1, 3,1, 1, 1, 2, 2, 2, 4, 2, 2, 2, 1, 1, 1, 1,0, 0, 2, 2,

3, 3,2, 2, 3,2, 0, 0, 1, 1,3, 3, 3, 1, 1, 1, 1, 1,2, 2, 2, 2, 1, 1, 1, 1, 0,1, 1, 1, 3,1, 1, 1, 2, 2, 2, 1,

1, 1,2, 1, 1, 1,3, 3,5, 3, 3, 1, 1, 1, 3, 3, 3, 3, 1, 1, 1,4, 1, 1, 4, 4, 4, 4, 4, 4,1, 1, 1,2, 2,5, 5, 2,

3, 3, 4, 4,3,2, 2, 2, 1,5, 1,2, 2, 1, 1, 1, 2, 2, 2, 2, 2,1, 1, 0,1, 1, 1,3, 3, 3, 3, 3

from scipy import stats

# Data

tickets = [1, 1, 0, 1, 2, 2, 0, 0, 0, 3, 3, 0, 3, 3, 0, 2, 2, 2, 1, 1, 4, 1, 1, 0, 3, 0,

           0, 0, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 4, 4, 4, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,

           1, 1, 1, 1, 1, 3, 3, 0, 3, 3, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 3, 3, 3, 2, 3, 3,

           1, 1, 1, 2, 2, 2, 4, 2, 2, 2, 1, 1, 1, 1, 0, 0, 2, 2, 3, 3, 2, 2, 3, 2, 0, 0,

           1, 1, 3, 3, 3, 1, 1, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 0, 1, 1, 1, 3, 1, 1, 1,

           2, 2, 2, 1, 1, 1, 2, 1, 1, 1, 3, 3, 5, 3, 3, 1, 1, 1, 3, 3, 3, 3, 1, 1, 1, 4,

           1, 1, 4, 4, 4, 4, 4, 4, 1, 1, 1, 2, 2, 5, 5, 2, 3, 3, 4, 4, 3, 2, 2, 2, 1, 5,

           1, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 1, 1, 0, 1, 1, 1, 3, 3, 3, 3, 3]

# Mean

mean\_tickets = np.mean(tickets)

# Median

median\_tickets = np.median(tickets)

# Mode

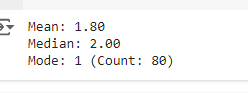
mode\_tickets = stats.mode(tickets)

print(f"Mean: {mean\_tickets:.2f}")

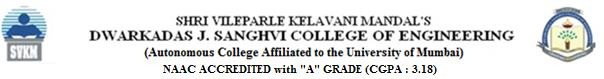
print(f"Median: {median\_tickets:.2f}")

# Access the mode and count directly from the ModeResult object

print(f"Mode: {mode\_tickets.mode} (Count: {mode\_tickets.count})")



1. Select any data set. (Kaggle, UCI Machine Learning Repository, Google data sets)
2. Generate the mean, median, mode, range, midrange for the chosen dataset.
3. Generate various types of graphs for the selected dataset.



# SOURCE CODE (OPTIONAL):

**OBSERVATIONS / DISCUSSION OF RESULT:**

# CONCLUSION:

**Observation Sheet Questions:**

1. Based on all conclusions on your actual results; describe the meaning of the experiment and the implications of your results.
2. Give some real-life examples, where these measures are applied.

# REFERENCES:

**Website References:**

Chapter 2 Lab 2: Descriptive Statistics | Answering questions with data: Lab Manual (crumplab.com)

<https://www.webpages.uidaho.edu/~stevel/251/Utts/R/chapter%2011.pdf> <https://onlinestatbook.com/2/estimation/mean.html>

https://sphweb.bumc.bu.edu/otlt/mph- modules/bs/bs704\_confidence\_intervals/bs704\_confidence\_intervals\_print.html

https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-2- sampling-distributions-and-confidence-intervals/