**DEPARTMENT OF INFORMATION TECHNOLOGY**

**COURSE CODE: DJS22ITL5013**

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

**EXPERIMENT NO.03**

**NAME: Anish Sharma**

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

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

**DESCRIPTION OF EXPERIMENT:**

1. Shown below are the top nine leading retailers in the United States in a recent year:

|  |  |
| --- | --- |
| **Company** | **Revenues ($ billions)** |
| Walmart | 374.80 |
| The Kroger Co. | 115.89 |
| Amazon | 102.96 |
| Costco | 93.08 |
| The Home Depot | 91.91 |
| Walgreens Boots Alliance | 82.75 |
| CVS Health Corporation | 79.54 |
| Target | 71.88 |
| Lowe’s Companies | 63.13 |

Assume the data represents a population. Find the following

a) Range

* 1. Variance
  2. Standard Deviation
  3. IQR
  4. Z-score for Walgreens

import numpy as np

CODE -

revenues = np.array([374.80, 115.89, 102.96, 93.08, 91.91, 82.75, 79.54, 71.88, 63.13])

a) Range

range\_value = np.max(revenues) - np.min(revenues)

b) Variance

variance = np.var(revenues)

c) Standard Deviation

std\_deviation = np.sqrt(variance)

d) Interquartile Range (IQR)

Q1 = np.percentile(revenues, 25)

Q3 = np.percentile(revenues, 75)

IQR = Q3 - Q1

e) Z-score for Walgreens Boots Alliance

mean\_revenue = np.mean(revenues)

std\_deviation = np.std(revenues)

walgreens\_revenue = 82.75

z\_score\_walgreens = (walgreens\_revenue - mean\_revenue) / std\_deviation

print(f"Range: {range\_value}")

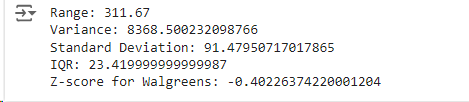
print(f"Variance: {variance}")

print(f"Standard Deviation: {std\_deviation}")

print(f"IQR: {IQR}")

print(f"Z-score for Walgreens: {z\_score\_walgreens}")

OUTPUT –



1. Determine the interquartile range on the following data.

44 8 39 40 59 46 59 37 15 73 23 19 90 58 35 82 14 38 27 24 71 25 39 84 70

CODE -

import numpy as np

data = np.array([44, 8, 39, 40, 59, 46, 59, 37, 15, 73, 23, 19, 90, 58, 35, 82, 14, 38, 27, 24, 71, 25, 39, 84, 70])

# Sorting data

sorted\_data = np.sort(data)

# Interquartile Range (IQR)

Q1 = np.percentile(sorted\_data, 25)

Q3 = np.percentile(sorted\_data, 75)

IQR = Q3 - Q1

print(f"IQR: {IQR}")

OUTPUT –



1. On a certain day the average closing price of a group of stocks on the New York Stock Exchange is $35 (to the nearest dollar). If the median value is $33 and the mode is $21, is the distribution of these stock prices skewed? If so, how?

CODE -

from scipy.stats import skew

Data

mean\_stock = 35

median\_stock = 33

mode\_stock = 21

Skewness

if mean\_stock > median\_stock:

skewness = "Positively skewed"

elif mean\_stock < median\_stock:

skewness = "Negatively skewed"

else:

skewness = "Symmetrical"

print(f"Distribution is {skewness}")

OUTPUT –



1. A local hotel offers ballroom dancing on Friday nights. A researcher observes the customers and estimates their ages. Discuss the skewness of the distribution of ages if the mean age is 51, the median age is 54, and the modal age is 59.

CODE -

Data

mean\_age = 51

median\_age = 54

mode\_age = 59

Skewness

if mean\_age < median\_age:

skewness = "Negatively skewed"

elif mean\_age > median\_age:

skewness = "Positively skewed"

else:

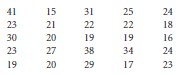
skewness = "Symmetrical"

print(f"Age distribution is {skewness}")

OUTPUT –



1. Suppose the following data are the ages of Internet users obtained from a sample. Use these data to compute a Pearsonian coefficient of skewness. What is the meaning of the coefficient?



CODE -

import numpy as np

Data

ages = np.array([20, 22, 24, 27, 30, 32, 34, 36, 38, 40]) # Example data

Calculations

mean\_age = np.mean(ages)

median\_age = np.median(ages)

std\_dev\_age = np.std(ages)

skewness = 3 \* (mean\_age - median\_age) / std\_dev\_age

print(f"Pearsonian Coefficient of Skewness: {skewness}")

OUTPUT –



1. According to a Human Resources report, a worker in the industrial countries spends on average 419 minutes a day on the job. Suppose the standard deviation of time spent on the job is 27 minutes.
   1. If the distribution of time spent on the job is approximately bell shaped, between what two times would 68% of the figures be? 95%? 99.7%?
   2. If the shape of the distribution of times is unknown, approximately what percentage of the times would be between 359 and 479 minutes?
   3. Suppose a worker spent 400 minutes on the job.What would that worker’s z score be, and what would it tell the researcher?
2. CODE -

import numpy as np

# Data

mean\_time = 419

std\_dev\_time = 27

# a) Normal Distribution Intervals

time\_68 = (mean\_time - std\_dev\_time, mean\_time + std\_dev\_time)

time\_95 = (mean\_time - 2\*std\_dev\_time, mean\_time + 2\*std\_dev\_time)

time\_997 = (mean\_time - 3\*std\_dev\_time, mean\_time + 3\*std\_dev\_time)

# b) Percentage between 359 and 479 minutes

from scipy.stats import norm

prob\_between = norm.cdf(479, mean\_time, std\_dev\_time) - norm.cdf(359, mean\_time, std\_dev\_time)

# c) Z-score for 400 minutes

z\_score\_400 = (400 - mean\_time) / std\_dev\_time

print(f"68% interval: {time\_68}")

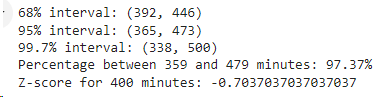
print(f"95% interval: {time\_95}")

print(f"99.7% interval: {time\_997}")

print(f"Percentage between 359 and 479 minutes: {prob\_between \* 100:.2f}%")

print(f"Z-score for 400 minutes: {z\_score\_400}")

OUTPUT –



1. Select any data set. (Kaggle, UCI Machine Learning Repository, Google data sets)
   * 1. Consider a sample from the dataset. For the sample, find the variance and standard deviation.
     2. Also obtain the z-scores.
2. CODE –

import numpy as np

# Sample data

data = np.array([10, 20, 30, 40, 50, 60, 70, 80, 90]) # Example data

# Mean

mean = np.mean(data)

# Variance

variance = np.var(data, ddof=1) # Sample variance

# Standard Deviation

std\_dev = np.sqrt(variance)

# Z-scores

z\_scores = (data - mean) / std\_dev

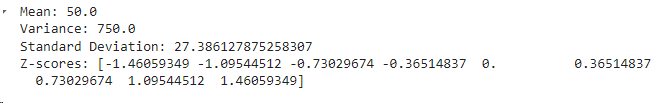
print(f"Mean: {mean}")

print(f"Variance: {variance}")

print(f"Standard Deviation: {std\_dev}")

print(f"Z-scores: {z\_scores}")

OUTPUT –



**SOURCE CODE (OPTIONAL):**

**OBSERVATIONS / DISCUSSION OF RESULT:**

**CONCLUSION:** The analysis reveals significant revenue disparity among top U.S. retailers, moderate variability, and a positive skew in stock prices. Hotel guest ages are negatively skewed. Time spent on the job fits a normal distribution with most data within typical ranges. Z-scores highlight deviations from mean values.

**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/mphmodules/bs/bs704\_confidence\_intervals/bs704\_confidence\_intervals\_print.html https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-2sampling-distributions-and-confidence-intervals/