Questions for Practice

1. Explain eigenvalues and eigenvectors. Discuss the limitations in eigenvalue decomposition and define singular value decomposition.

2. Determine the singular value decomposition (SVD) of the matrix A, expressed as

$$A = U\Sigma V^T$$

   where

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}.$$

3. Define principal component analysis. Discuss the application in singular value decomposition in Principal Component Analysis.

4. What is R software and discuss its applications. Explain the features of *RStudio*. Provide a brief discussion on *command line* and *scripts* in *R*.

5. For the given DataFrame 'df', complete the Python code appropriately:

```
import pandas as pd
import numpy as np
data = {
    'Name': ['Rohan', 'Shreya', 'Kavya', 'Vansh', 'Kushal'],
    'Age': [25, 30, 35, np.nan, 40],
    'Salary': [50000, 60000, np.nan, 80000, 90000],
    'Department': ['HR', 'IT', 'IT', 'Finance', 'HR']}
df = pd.DataFrame(data)
```

   Based on above, complete the following sentences:

   (a) To fill missing values in the "Age" column with the mean age, we use ....

   (b) To drop all rows that contain any NaN values, we use ....

   (c) To filter the DataFrame to show only employees from the "IT" department, we use ....

   (d) To get the total number of missing values in the entire DataFrame, we use ....

Questions for Practice

1. Define descriptive and inferential Statistics. Discuss the measure of central tendency and its applications.

2. Explain the problems of parameter estimation and hypothesis testing in Statistics.

3. Define population, sample, and discuss the need for random sampling in Statistics.

4. Write the commands for generating random sample using R software for

   - Poisson Distribution
   - Normal Distribution
   - Geometric Distribution
   - Binomial Distribution
   - Uniform Distribution

   Further, generate the histogram in each case. Demonstrate the plotting of bar graph, pie chart, line plot, and box plot using R software.

5. For the given DataFrame 'df', complete the Python code appropriately:

```
import pandas as pd
import numpy as np
data = {
    'Name': ['Rohan', 'Shreya', 'Kavya', 'Vansh', 'Kushal'],
    'Age': [25, 30, 35, np.nan, 40],
    'Salary': [50000, 60000, np.nan, 80000, 90000],
    'Department': ['HR', 'IT', 'IT', 'Finance', 'HR']}
df = pd.DataFrame(data)
```

based on above, complete the following sentences:

   (a) To create a new column "Experience" with random values between 1 and 10, we use ....

   (b) To sort the DataFrame by "Salary" in descending order, we use .... (by='Salary', ascending=False).

   (c) To select only the columns "Name" and "Salary", we use ....

   (d) To check for duplicate rows in the DataFrame, we use ....