

In [30]:

```
1 import pandas as pd
2 df = pd.read_csv(r"C:\Users\Anusha V\Desktop\student data.csv")
3 df
```

Out[30]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24	25	26	27	28	29	30	COURSE ID
0	STUDENT1	2	2	3	3	1	2	2	1	1	...	1	1	3	2	1	2	1	1	
1	STUDENT2	2	2	1	3	1	2	2	1	1	...	1	1	3	2	3	2	2	3	
2	STUDENT3	2	2	2	3	2	2	2	2	4	...	1	1	2	2	1	1	2	2	
3	STUDENT4	1	1	1	3	1	2	1	2	1	...	1	2	3	2	2	1	3	2	
4	STUDENT5	2	2	1	3	2	2	1	3	1	...	2	1	2	2	2	1	2	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
140	STUDENT141	2	1	2	3	1	1	2	1	1	...	1	1	2	1	2	1	3	3	
141	STUDENT142	1	1	2	4	2	2	2	1	4	...	1	1	3	2	2	1	5	3	
142	STUDENT143	1	1	1	4	2	2	2	1	1	...	1	1	3	3	2	1	4	3	
143	STUDENT144	2	1	2	4	1	1	1	5	2	...	2	1	2	1	2	1	5	3	
144	STUDENT145	1	1	1	5	2	2	2	3	1	...	2	1	3	2	3	1	5	4	

145 rows × 33 columns

In [31]:

```
1 df.tail()
```

Out[31]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24	25	26	27	28	29	30	COURSE ID
140	STUDENT141	2	1	2	3	1	1	2	1	1	...	1	1	2	1	2	1	3	3	9
141	STUDENT142	1	1	2	4	2	2	2	1	4	...	1	1	3	2	2	1	5	3	9
142	STUDENT143	1	1	1	4	2	2	2	1	1	...	1	1	3	3	2	1	4	3	9
143	STUDENT144	2	1	2	4	1	1	1	5	2	...	2	1	2	1	2	1	5	3	9
144	STUDENT145	1	1	1	5	2	2	2	3	1	...	2	1	3	2	3	1	5	4	9

5 rows × 33 columns

In [3]:

1df.head()

Out[3]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24	25	26	27	28	29	30	COURSE ID	C
0	STUDENT1	2	2	3.0	3	1	2	2	1	1	...	1	1	3	2	1	2	1	1	1	
1	STUDENT2	2	2	NaN	3	1	2	2	1	1	...	1	1	3	2	3	2	2	3	1	
2	STUDENT3	2	2	2.0	3	2	2	2	2	4	...	1	1	2	2	1	1	2	2	1	
3	STUDENT4	1	1	1.0	3	1	2	1	2	1	...	1	2	3	2	2	1	3	2	1	
4	STUDENT5	2	2	NaN	3	2	2	1	3	1	...	2	1	2	2	2	1	2	2	1	

5 rows × 33 columns

In [4]:

1df.isnull()

Out[4]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24
0		False	False	False	False	False	False	False	False	False	...	False	False
1		False	False	False	True	False	False	False	False	False	...	False	False
2		False	False	False	False	False	False	False	False	False	...	False	False
3		False	False	False	False	False	False	False	False	False	...	False	False
4		False	False	False	True	False	False	False	False	False	...	False	False
...		...	...	...	...	...	...	...	...	...	...	...	...
140		False	False	False	False	False	False	False	False	False	...	False	False
141		False	False	False	False	False	False	False	False	False	...	False	False
142		False	False	False	False	False	False	False	False	False	...	False	False
143		False	False	False	False	False	False	False	False	False	...	False	False
144		False	False	False	False	False	False	False	False	False	...	False	False

145 rows × 33 columns

In [5]:

1df.fillna(1)

Out[5]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24	25	26	27	28	29	30	COL
0	STUDENT1	2	2	3.0	3	1	2	2	1	1	...	1	1	3	2	1	2	1	1	
1	STUDENT2	2	2	1.0	3	1	2	2	1	1	...	1	1	3	2	3	2	2	3	
2	STUDENT3	2	2	2.0	3	2	2	2	2	4	...	1	1	2	2	1	1	2	2	
3	STUDENT4	1	1	1.0	3	1	2	1	2	1	...	1	2	3	2	2	1	3	2	
4	STUDENT5	2	2	1.0	3	2	2	1	3	1	...	2	1	2	2	2	1	2	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
140	STUDENT141	2	1	2.0	3	1	1	2	1	1	...	1	1	2	1	2	1	3	3	
141	STUDENT142	1	1	2.0	4	2	2	2	1	4	...	1	1	3	2	2	1	5	3	
142	STUDENT143	1	1	1.0	4	2	2	2	1	1	...	1	1	3	3	2	1	4	3	
143	STUDENT144	2	1	2.0	4	1	1	1	5	2	...	2	1	2	1	2	1	5	3	
144	STUDENT145	1	1	1.0	5	2	2	2	3	1	...	2	1	3	2	3	1	5	4	

145 rows × 33 columns

In [6]:

1df.dropna(0)

C:\Users\Anusha V\AppData\Local\Temp\ipykernel\_9948\4129559070.py:1: FutureWarning: In a future version of pandas all arguments of DataFrame.dropna will be keyword-only.  
df.dropna(0)

Out[6]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	23	24	25	26	27	28	29	30	COL
0	STUDENT1	2	2	3.0	3	1	2	2	1	1	...	1	1	3	2	1	2	1	1	
2	STUDENT3	2	2	2.0	3	2	2	2	2	4	...	1	1	2	2	1	1	2	2	
3	STUDENT4	1	1	1.0	3	1	2	1	2	1	...	1	2	3	2	2	1	3	2	
5	STUDENT6	2	2	2.0	3	2	2	2	2	1	...	1	1	1	2	1	2	4	4	
6	STUDENT7	1	2	1.0	4	2	2	2	1	1	...	1	1	3	3	3	3	4	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
140	STUDENT141	2	1	2.0	3	1	1	2	1	1	...	1	1	2	1	2	1	3	3	
141	STUDENT142	1	1	2.0	4	2	2	2	1	4	...	1	1	3	2	2	1	5	3	
142	STUDENT143	1	1	1.0	4	2	2	2	1	1	...	1	1	3	3	2	1	4	3	
143	STUDENT144	2	1	2.0	4	1	1	1	5	2	...	2	1	2	1	2	1	5	3	
144	STUDENT145	1	1	1.0	5	2	2	2	3	1	...	2	1	3	2	3	1	5	4	

143 rows × 33 columns

```
In [9]: 1 df = pd.read_csv(r"C:\Users\Anusha V\Desktop\student data.csv")
        2 df.isnull().sum()
```

```
Out[9]: STUDENT ID      0
        1              0
        2              0
        3              0
        4              0
        5              0
        6              0
        7              0
        8              0
        9              0
       10              0
       11              0
       12              0
       13              0
       14              0
       15              0
       16              0
       17              0
       18              0
       19              0
       20              0
       21              0
       22              0
       23              0
       24              0
       25              0
       26              0
       27              0
       28              0
       29              0
       30              0
      COURSE ID      0
      GRADE          0
      dtype: int64
```

```
In [10]: 1 df.isnull().sum().sum()
```

```
Out[10]: 0
```

```
In [11]: 1 import pandas as pd
2 from scipy import stats
3 column_name = "1"
4 df['z-score'] = stats.zscore(df['2'])
5 df
```

Out[11]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	24	25	26	27	28	29	30	COURSE ID
0	STUDENT1	2	2	3	3	1	2	2	1	1	...	1	3	2	1	2	1	1	1
1	STUDENT2	2	2	1	3	1	2	2	1	1	...	1	3	2	3	2	2	3	1
2	STUDENT3	2	2	2	3	2	2	2	2	4	...	1	2	2	1	1	2	2	1
3	STUDENT4	1	1	1	3	1	2	1	2	1	...	2	3	2	2	1	3	2	1
4	STUDENT5	2	2	1	3	2	2	1	3	1	...	1	2	2	2	1	2	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
140	STUDENT141	2	1	2	3	1	1	2	1	1	...	1	2	1	2	1	3	3	9
141	STUDENT142	1	1	2	4	2	2	2	1	4	...	1	3	2	2	1	5	3	9
142	STUDENT143	1	1	1	4	2	2	2	1	1	...	1	3	3	2	1	4	3	9
143	STUDENT144	2	1	2	4	1	1	1	5	2	...	1	2	1	2	1	5	3	9
144	STUDENT145	1	1	1	5	2	2	2	3	1	...	1	3	2	3	1	5	4	9

145 rows × 34 columns

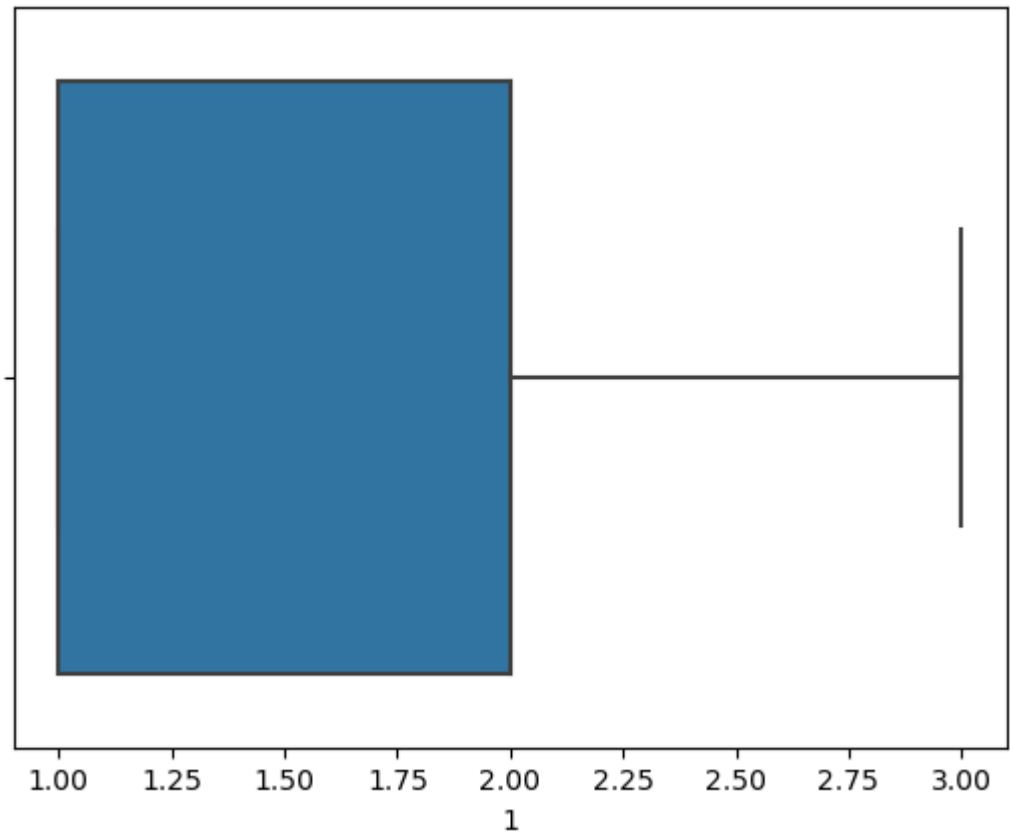
```
In [12]: 1 import pandas as pd
2 Q1 = df['1'].quantile(0.2)
3 Q3 = df['2'].quantile(0.20)
4 IQR = Q1 - Q3
5 lower_bound = Q1 - 1.5*IQR
6 upper_bound = Q3 + 1.5*IQR
7 outliers = [(df['1'] < lower_bound) | (df['2'] > upper_bound)]
8 outliers
```

Out[12]:

0	True
1	True
2	True
3	False
4	True
...	...
140	False
141	False
142	False
143	False
144	False

Length: 145, dtype: bool]

```
In [15]: 1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 sns.boxplot(x=df['1'])
5 plt.show()
```



```
In [16]: 1 df.head()
```

Out[16]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	24	25	26	27	28	29	30	COURSE ID	GRADE
0	STUDENT1	2	2	3	3	1	2	2	1	1	...	1	3	2	1	2	1	1	1	1
1	STUDENT2	2	2	1	3	1	2	2	1	1	...	1	3	2	3	2	2	3	1	1
2	STUDENT3	2	2	2	3	2	2	2	2	4	...	1	2	2	1	1	2	2	1	1
3	STUDENT4	1	1	1	3	1	2	1	2	1	...	2	3	2	2	1	3	2	1	1
4	STUDENT5	2	2	1	3	2	2	1	3	1	...	1	2	2	2	1	2	2	1	1

5 rows × 34 columns

In [18]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 df['subject'] = df['COURSE ID']*2
4 df
```

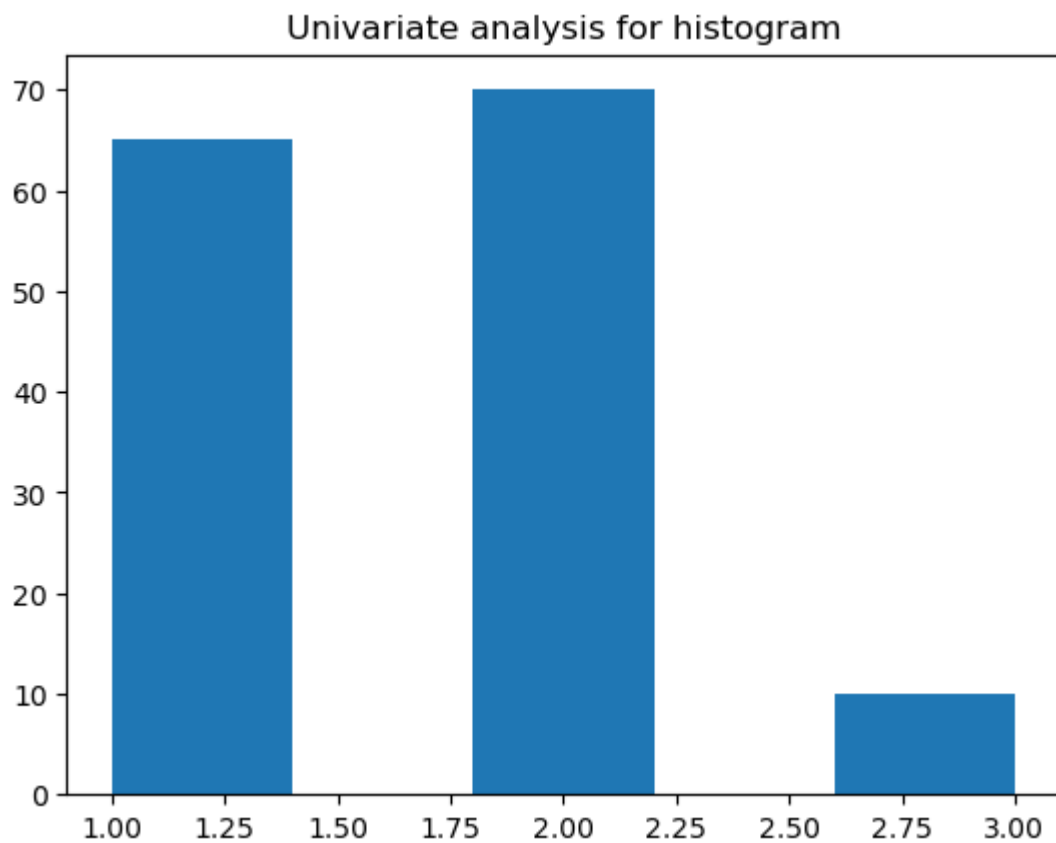
Out[18]:

	STUDENT ID	1	2	3	4	5	6	7	8	9	...	25	26	27	28	29	30	COURSE ID	GR
0	STUDENT1	2	2	3	3	1	2	2	1	1	...	3	2	1	2	1	1	1	
1	STUDENT2	2	2	1	3	1	2	2	1	1	...	3	2	3	2	2	3	1	
2	STUDENT3	2	2	2	3	2	2	2	2	4	...	2	2	1	1	2	2	1	
3	STUDENT4	1	1	1	3	1	2	1	2	1	...	3	2	2	1	3	2	1	
4	STUDENT5	2	2	1	3	2	2	1	3	1	...	2	2	2	1	2	2	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
140	STUDENT141	2	1	2	3	1	1	2	1	1	...	2	1	2	1	3	3	9	
141	STUDENT142	1	1	2	4	2	2	2	1	4	...	3	2	2	1	5	3	9	
142	STUDENT143	1	1	1	4	2	2	2	1	1	...	3	3	2	1	4	3	9	
143	STUDENT144	2	1	2	4	1	1	1	5	2	...	2	1	2	1	5	3	9	
144	STUDENT145	1	1	1	5	2	2	2	3	1	...	3	2	3	1	5	4	9	

145 rows × 35 columns

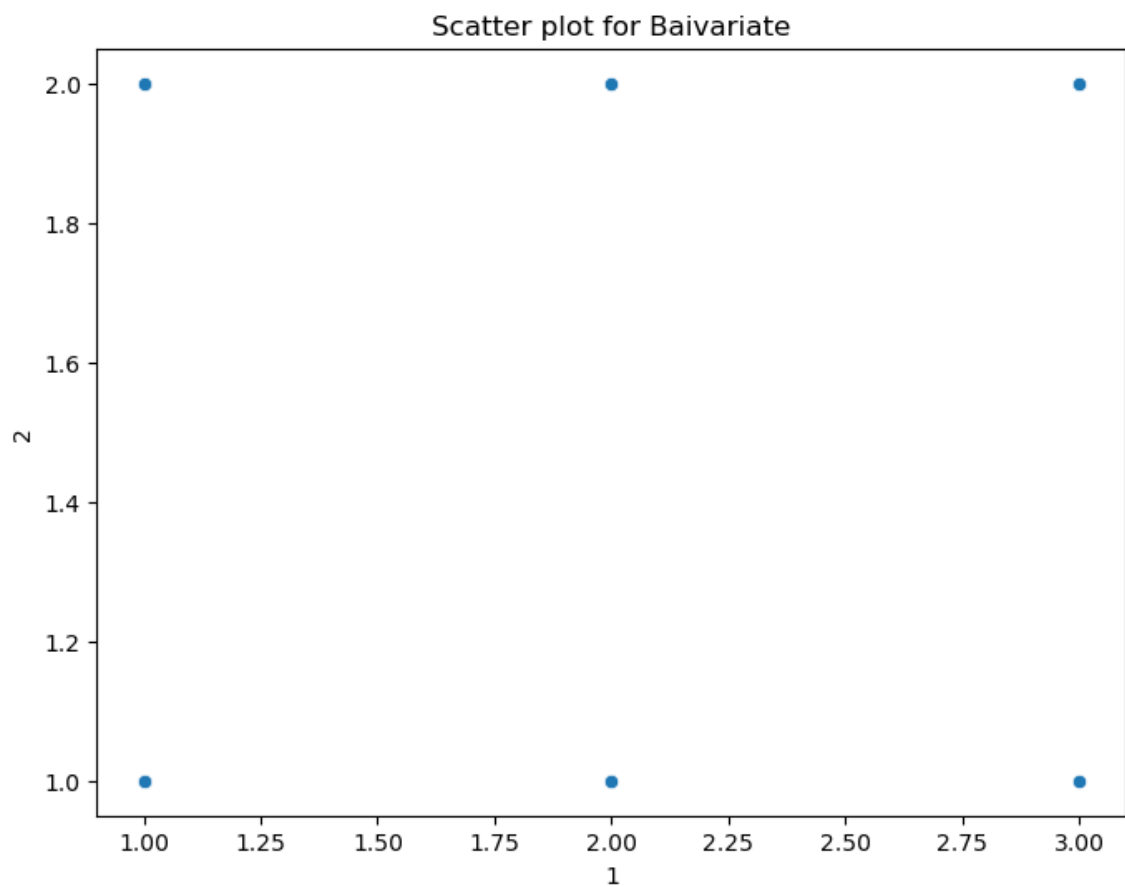
```
In [24]: 1 import pandas as pd
2 column_name = ("1")
3 mean = df['1'].mean()
4 median = df['1'].median()
5 mode = df['1'].mode().iloc[0]
6 print("mean:{'mean'}")
7 print("median:{median}")
8 print("mode:{mode}")
9 plt.hist(df["1"], bins=5)
10 plt.title("Univariate analysis for histogram")
11 plt.show()
```

```
mean:{'mean'}
median:{median}
mode:{mode}
```





```
In [29]: 1 import seaborn as sns
2 column1 = df['1']
3 column2 = df['2']
4 plt.figure(figsize=(8, 6))
5 sns.scatterplot(x=column1, y=column2)
6 plt.title("Scatter plot for Baivariate")
7 plt.show()
```



```
In [ ]: 1
```