

# Descriptive Statistics

## Assignment

### Background:

The data for analysis is an insurance sector data in which premiums information is provided for each policy holder for all the regions and zones.

### Questions

- A. Import Premiums data.
- B. Obtain the Mode(max count) for the count of policies available across each Zone.
- C. Obtain box-whisker plots for Vintage period. Detect outliers if present.
- D. Find skewness and kurtosis of Premium amount by Zone.
- E. Draw a scatter plot of Premium and Vintage period. Find the correlation coefficient between Premium and Vintage period and interpret the value.

### Solutions

#Q1. Import Premiums data

##A.

```
import pandas as pd
premium = pd.read_csv("Premiums.csv")
premium.head()
```

#Q2. Obtain the Mode for the count of policies available across each Zone

##A.

```
from scipy import stats
```

```
freq = premium['ZONE_NAME'].value_counts()
freq
```

#Interpretation:Mode is 2634 for South Zone

#Q3. Obtain box-whisker plots for Vintage period in each zone. Detect and remove outliers if present. Hint: use Boxplot() function of 'car' Package

##A.

```
import matplotlib.pyplot as plt
premium.Vintage_Period.plot.box(patch_artist=True,color="darkorange",label="");plt.title( "BoxPlot (PREMIUM)");plt.ylabel("Vintage Period")
```

#Q4. Find skewness and kurtosis of Premium amount by Zone.

##A.

```
premium.groupby('ZONE_NAME')['Premium'].skew()
premium.groupby('ZONE_NAME')['Premium'].apply(pd.DataFrame.kurt)
```

#Q5. Find the correlation coefficient between Premium and Vintage period and interpret the value.

##A.

```
plt.scatter(premium.Premium,premium.Vintage_Period, color='red')
```

```
import numpy as np
```

```
np.corrcoef(premium.Premium,premium.Vintage_Period)
```

#Interpretation: There is positive relation between premium and vintage period but the relation is of less value