

Descriptive Statistics

Assignment 1

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-

1. Import Premiums data.
2. Obtain the Mode for the count of policies available across each Zone.
3. Obtain box-whisker plots for Vintage period.
Detect outliers if present. Hint: use Boxplot() function of 'car' Package
4. Find skewness and kurtosis of Premium amount by Zone.
5. Draw a scatter plot of Premium and Vintage period.
6. Find the correlation coefficient between Premium and Vintage period and interpret the value.

Descriptive Statistics

Assignment Solution Sample 1

```
#Q1. Import Premiums data
##A.
premium<-read.csv(file.choose(),header=TRUE)
head(premium)
```

```
#Q2. Obtain the Mode for the count of policies available across each Zone
##A.
freq <- table(premium$ZONE_NAME)
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.
library(car)
Boxplot(premium$Vintage_Period, data= premium, main = "BoxPlot
(PREMIUM)",ylab = "Vintage Period",col = "darkorange")
```

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

##A.

```
library(e1071)
```

```
f <- function(x){skew = skewness(x,type = 2),kurt = kurtosis(x,type = 2)}
```

```
aggregate(Premium~ZONE_NAME,data = premium,FUN = f)
```

#Q5. Draw a scatter plot of Premium and Vintage period. Find the correlation coefficient between Premium and Vintage period and interpret the value.

##A.

```
plot(premium$Premium,premium$Vintage_Period,col="red")
```

```
cor(premium$Premium,premium$Vintage_Period)
```

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