Question 1

Perform an analysis of variance to determine whether differences in PlantGrowth (as measured by weight of batch) between the three groups are statistically significant. Use PlantGrowth dataset in R.

```
#perform the ANOVA and save the results to plantanova
plantanova<-aov(weight~group, PlantGrowth)

#view the ANOVA data
anova(plantanova)

## Analysis of Variance Table
##
## Response: weight
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## group      2  3.7663  1.8832  4.8461 0.01591 *
## Residuals 27 10.4921  0.3886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#p-value = 0.0159. p-value is less than the significance level of 0.05, So we
can reject the null hypothesis that the mean growth is the same for all
treatments.

#p-value = 0.0159.This means that, if the effect of all three treatments were
the same, we would have less than 2% chance of seeing differences between
groups as large or larger than this

coef(plantanova)

## (Intercept)   grouptrt1   grouptrt2
##       5.032      -0.371       0.494

#Control treatment gives an average weight of 5.032. The effect of treatment
1 (trt1) is to reduce weight by an average of -0.371 units compared to the
control method, and the effect of treatment 2 (trt2) is to increase weight by
an average of 0.494 units compared to the control method.
```

Question 2

The dataset gives data for a sample of 20 bottles of soft drink taken from a filling line. The dataset contains one variable named Volume, which gives the volume of liquid in millilitres for each of the bottles.

The bottle filling volume is believed to follow a normal distribution with a mean of 500 milliliters and a standard deviation of 25 milliliters. Suppose that you wish to use a one-sample Kolmogorov-Smirnov test to determine whether the data is consistent with this theory. The test has the null hypothesis that the bottles volumes are drawn from the described distribution, and the alternative hypothesis that they are not. A significance level of 0.05 will be used for the test.

Check for any evidence that the bottle volumes are not drawn from the described normal distribution.

```
bottles<-read.csv("C:/Ravi/R Topgear/Part 2/bottles.csv")
ks.test(bottles$Volume, "pnorm", 500, 25)

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  bottles$Volume
## D = 0.22879, p-value = 0.2108
## alternative hypothesis: two-sided

#The p-value for the one-sample Kolmogorov-Smirnov test is 0.2108. As this is
not less than the significance level of 0.05, we can not reject the null
hypothesis. This means that there is no evidence that the bottle volumes are
not drawn from the normal distribution.
```

Question 3

Suppose that a fair die is rolled 10 times. What is the probability of throwing exactly two sixes?

```
dbinom(2,10,1/6)

## [1] 0.29071

#The probability of throwing exactly two sixes, when a fair die is rolled 10
times is 0.29071 or 29.071%
```

Question 4

The number of lobster ordered in a restaurant on a given day is known to follow a Poisson distribution with a mean of 20. What is the probability that exactly eighteen lobsters will be ordered tomorrow?

```
dpois(18,20)

## [1] 0.08439355
```

```
# the probability that exactly eighteen lobsters will be ordered tomorrow is
8.4%
```

Question 5

using scale function normalize airquality dataset using Min−max normalization

```
airquality_dataset<-as.data.frame(airquality)
summary(airquality_dataset)

##      Ozone           Solar.R           Wind             Temp      
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
##  NA's   :37       NA's   :7                                       
##      Month            Day      
##  Min.   :5.000   Min.   : 1.0  
##  1st Qu.:6.000   1st Qu.: 8.0  
##  Median :7.000   Median :16.0  
##  Mean   :6.993   Mean   :15.8  
##  3rd Qu.:8.000   3rd Qu.:23.0  
##  Max.   :9.000   Max.   :31.0  
## 

airquality_dataset=na.omit(airquality_dataset)

max_data<-apply(airquality_dataset,2,max)
min_data<-apply(airquality_dataset,2, min)
data_scaled<-scale(airquality_dataset,center=min_data,scale=max_data-
min_data)
summary(data_scaled)

##      Ozone           Solar.R            Wind             Temp       
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.1018   1st Qu.:0.3257   1st Qu.:0.2772   1st Qu.:0.3500  
##  Median :0.1796   Median :0.6116   Median :0.4022   Median :0.5500  
##  Mean   :0.2461   Mean   :0.5437   Mean   :0.4152   Mean   :0.5198  
##  3rd Qu.:0.3653   3rd Qu.:0.7599   3rd Qu.:0.5000   3rd Qu.:0.6875  
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
##      Month             Day        
##  Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.2500   1st Qu.:0.2667  
##  Median :0.5000   Median :0.5000  
##  Mean   :0.5541   Mean   :0.4982  
##  3rd Qu.:1.0000   3rd Qu.:0.7167  
##  Max.   :1.0000   Max.   :1.0000
```

Question 6

using scale function normalize airquality dataset using Z-score Standardization

```
airquality_dataset<-as.data.frame(airquality)
summary(airquality_dataset)
```

```
##      Ozone           Solar.R           Wind              Temp      
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
##  NA's   :37       NA's   :7                                       
##      Month            Day      
##  Min.   :5.000   Min.   : 1.0  
##  1st Qu.:6.000   1st Qu.: 8.0  
##  Median :7.000   Median :16.0  
##  Mean   :6.993   Mean   :15.8  
##  3rd Qu.:8.000   3rd Qu.:23.0  
##  Max.   :9.000   Max.   :31.0  
## 
```

```
airquality_dataset=na.omit(airquality_dataset)

mean_data<-apply(airquality_dataset, 2, mean)
std_data<-apply(airquality_dataset, 2, sd)

data_scaled<-scale(airquality_dataset, center=mean_data, scale=std_data)
summary(data_scaled)
```

```
##      Ozone             Solar.R             Wind               Temp        
##  Min.   :-1.2351   Min.   :-1.9506   Min.   :-2.14735   Min.   :-2.1818  
##  1st Qu.:-0.7242   1st Qu.:-0.7822   1st Qu.:-0.71384   1st Qu.:-0.7128  
##  Median :-0.3335   Median : 0.2435   Median :-0.06736   Median : 0.1267  
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000  
##  3rd Qu.: 0.5981   3rd Qu.: 0.7756   3rd Qu.: 0.43859   3rd Qu.: 0.7038  
##  Max.   : 3.7835   Max.   : 1.6368   Max.   : 3.02452   Max.   : 2.0155  
##      Month              Day           
##  Min.   :-1.5041   Min.   :-1.716505  
##  1st Qu.:-0.8254   1st Qu.:-0.797725  
##  Median :-0.1467   Median : 0.006208  
##  Mean   : 0.0000   Mean   : 0.000000  
##  3rd Qu.: 1.2106   3rd Qu.: 0.752717  
##  Max.   : 1.2106   Max.   : 1.728921
```