

DA 1
PMDS503P - Statistical Inference Lab

Name : Anish Kumar Barik
RegNo: 24MDT0170

```
#Question No: 1

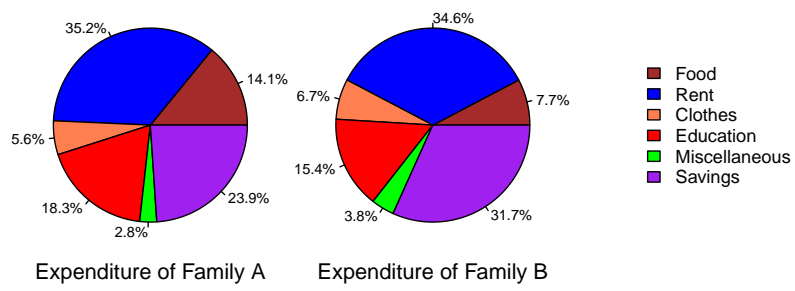
typeofcommodity <- c('Food', 'Rent', 'Clothes', 'Education', 'Miscellaneous', 'Savings')
familyA <- c(10, 25, 4, 13, 2, 17)
familyB <- c(8, 36, 7, 16, 4, 33)
data <- data.frame(typeofcommodity, familyA, familyB)

familyA_labels <- paste0(round(familyA / sum(familyA) * 100, 1), "%")
familyB_labels <- paste0(round(familyB / sum(familyB) * 100, 1), "%")

par(mfrow = c(1, 3), mar = c(2, 2, 2, 2), oma = c(2, 2, 2, 2), pin = c(2, 2))
colors <- c('brown', 'blue', 'coral', 'red', 'green', 'purple')
pie(familyA, labels = familyA_labels, col = colors)
mtext("Expenditure of Family A", side = 1, line = 0.5, cex = 1)

pie(familyB, labels = familyB_labels, col = colors)
mtext("Expenditure of Family B", side = 1, line = 0.5, cex = 1)

plot.new()
legend("center", legend = typeofcommodity, cex = 1.2, bty = "n", fill = colors)
```



#Question No: 2

`datasets::npk`

```
##      block N P K yield
## 1      1 0 1 1  49.5
## 2      1 1 1 0  62.8
## 3      1 0 0 0  46.8
## 4      1 1 0 1  57.0
## 5      2 1 0 0  59.8
## 6      2 1 1 1  58.5
## 7      2 0 0 1  55.5
## 8      2 0 1 0  56.0
## 9      3 0 1 0  62.8
## 10     3 1 1 1  55.8
## 11     3 1 0 0  69.5
## 12     3 0 0 1  55.0
## 13     4 1 0 0  62.0
## 14     4 1 1 1  48.8
## 15     4 0 0 1  45.5
## 16     4 0 1 0  44.2
## 17     5 1 1 0  52.0
## 18     5 0 0 0  51.5
## 19     5 1 0 1  49.8
## 20     5 0 1 1  48.8
## 21     6 1 0 1  57.2
## 22     6 1 1 0  59.0
## 23     6 0 1 1  53.2
## 24     6 0 0 0  56.0
```

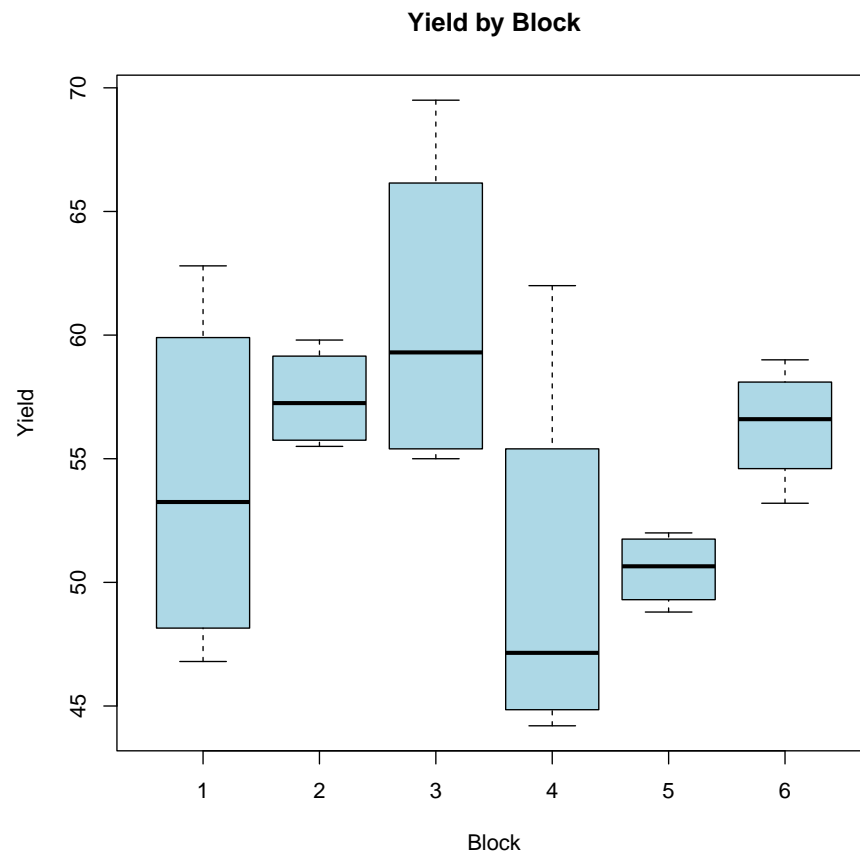
#The number of variables in dataset

`ncol(npk)`

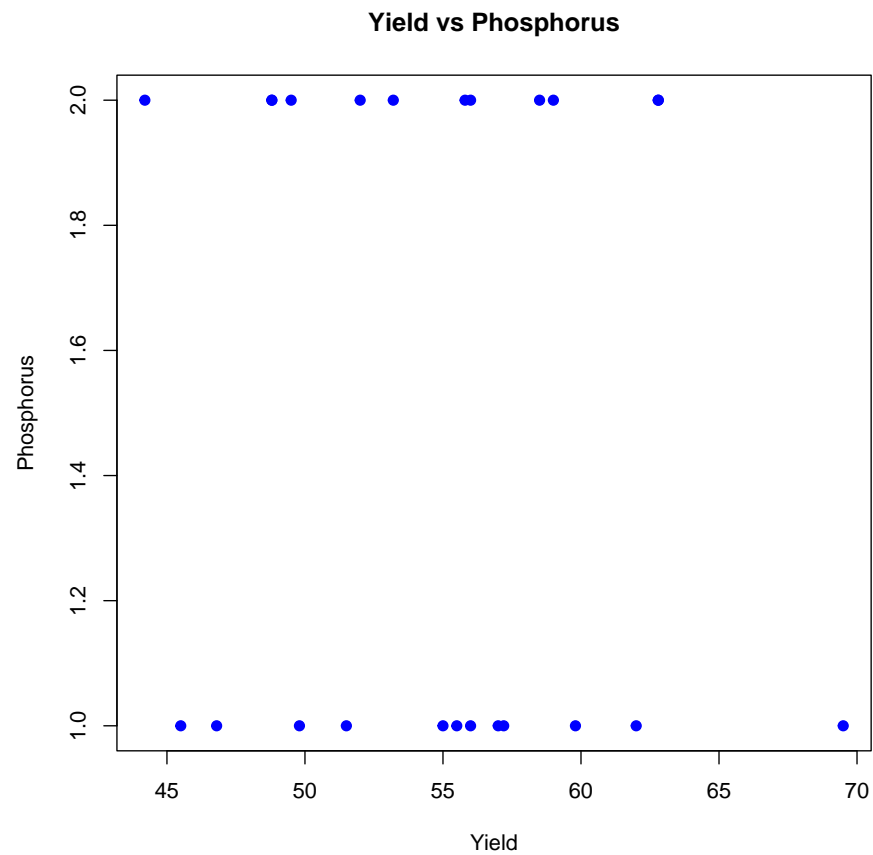
```
## [1] 5
```

#a box plot for any two variables.

`boxplot(yield ~ block, data=npk, main="Yield by Block", xlab="Block", ylab="Yield", col="lightblue")`

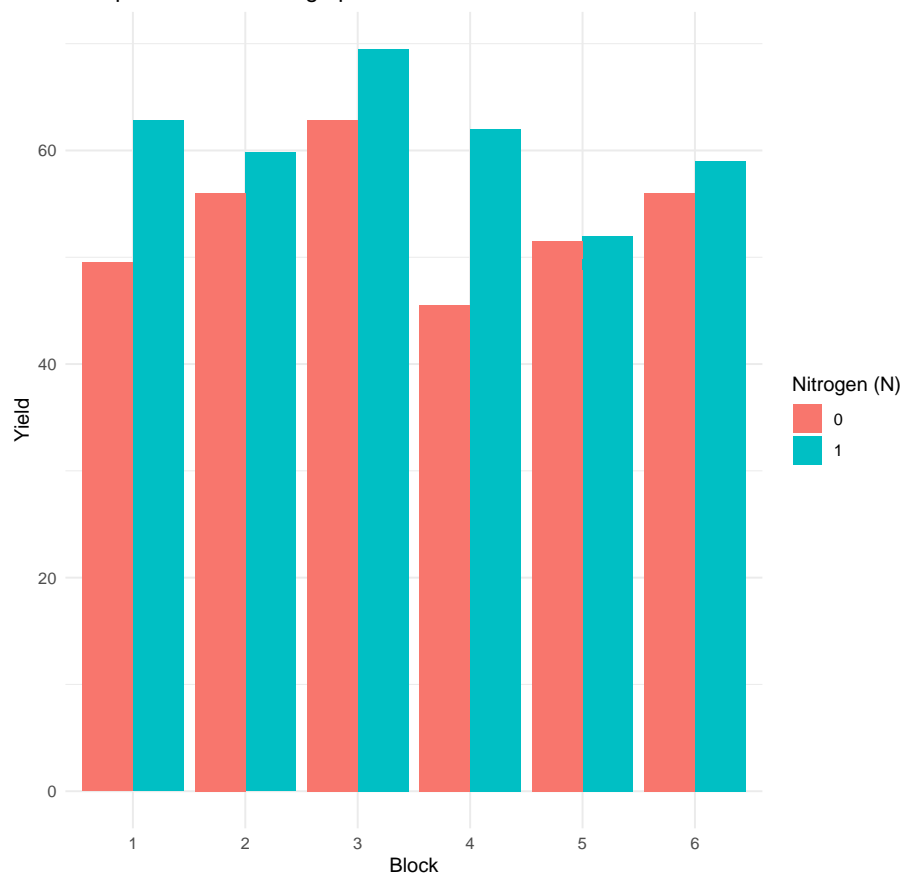


```
#Scatterplot for any two variables  
plot(npk$yield, npk$P, main="Yield vs Phosphorus", xlab="Yield",ylab="Phosphorus",col="blue")
```



```
#multiple bar chart
library(ggplot2)
npk$block <- as.factor(npk$block)
ggplot(npk, aes(x = block, y = yield, fill = factor(N))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Multiple Bar Chart using npk Dataset", x = "Block", y = "Yield", fill = "Nitrogen")
```

Multiple Bar Chart using npk Dataset



```
#summary
summary(npk)

## block N      P      K      yield
## 1:4 0:12 0:12 0:12 Min.   :44.20
## 2:4 1:12 1:12 1:12 1st Qu.:49.73
## 3:4                      Median :55.65
## 4:4                      Mean   :54.88
## 5:4                      3rd Qu.:58.62
## 6:4                      Max.   :69.50
```

```

#Question No: 3
library(readxl)
marks<- read_excel("C:/Users/sreej/Downloads/marks.xlsx")
marks

## # A tibble: 68 x 7
##   'Sl No' CAT1 CAT2 DA FAT QUIZ1 QUIZ2
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      1      27      36      10      81      16      14
## 2      2      26      36      10      73      16      12
## 3      3      10      28      10      50      12      14
## 4      4       5       0      10       0      18      12
## 5      5      43      48      10      94      12      16
## 6      6      25      25      10      61      12      14
## 7      7      29      35      10      81      14      20
## 8      8      28      49      10      80      14      14
## 9      9      35      46      10      86      20      16
## 10     10     11      10      10      20      15      15
## # i 58 more rows

summary(marks)

##      Sl No      CAT1      CAT2      DA      FAT
## Min.   : 1.00   Min.   : 5.00   Min.   : 0.00   Min.   :10   Min.   : 0.00
## 1st Qu.:17.75   1st Qu.:15.00   1st Qu.:25.00   1st Qu.:10   1st Qu.:49.00
## Median :34.50   Median :21.00   Median :34.00   Median :10   Median :63.00
## Mean   :34.50   Mean   :22.51   Mean   :31.62   Mean   :10   Mean   :61.25
## 3rd Qu.:51.25   3rd Qu.:28.00   3rd Qu.:39.25   3rd Qu.:10   3rd Qu.:80.00
## Max.   :68.00   Max.   :45.00   Max.   :49.00   Max.   :10   Max.   :94.00
##      QUIZ1      QUIZ2
## Min.    : 6.00   Min.    : 0.00
## 1st Qu.:12.00   1st Qu.:12.00
## Median :12.00   Median :14.00
## Mean    :12.82   Mean    :13.74
## 3rd Qu.:14.00   3rd Qu.:16.00
## Max.    :20.00   Max.    :20.00

sd(marks$CAT1)

## [1] 9.487608

sd(marks$CAT2)

## [1] 10.71062

sd(marks$QUIZ1)

## [1] 2.890753

```

```
sd(marks$QUIZ2)
## [1] 3.308143
sd(marks$DA)
## [1] 0
sd(marks$FAT)
## [1] 21.15005
library(e1071)
skewness(marks$CAT1 )
## [1] 0.4986285
skewness(marks$CAT2)
## [1] -0.516568
skewness(marks$DA)
## [1] NaN
skewness(marks$QUIZ1)
## [1] 0.2537272
skewness(marks$QUIZ2)
## [1] -0.6330455
skewness(marks$FAT)
## [1] -0.5308485
kurtosis(marks$CAT1 )
## [1] -0.2310864
kurtosis(marks$CAT2)
## [1] -0.1810742
kurtosis(marks$DA)
## [1] NaN
kurtosis(marks$QUIZ1)
## [1] 0.6760371
kurtosis(marks$QUIZ2)
## [1] 3.227198
kurtosis(marks$FAT)
## [1] -0.2490257
```


Descriptive Statistics Table						
Measure	CAT 1	CAT 2	DA	QUIZ 1	QUIZ 2	FAT
Mean	22.51	31.62	10	12.82	13.74	61.25
Median	21	34	10	12	14	63
Standard Deviation	9.49	10.71	0	2.89	3.31	21.15
Skewness	0.499	-0.517	NaN	0.254	-0.633	-0.531
Kurtosis	-0.231	-0.181	NaN	0.676	3.227	-0.249
1st Quartile	15.00	25.00	10	12	12	49
3rd Quartile	28	39.25	10	14	16	80