Assignment2

Pratishtha Deep

October 11, 2017

```
#Que:1 Create two matrices from the given set of numbers X1 =
{2,3,7,1,6,2,3,5,1} and x2 = {3,2,9,0,7,8,5,8,2}
#Answer:
A = matrix(c(2,3,7,1,6,2,3,5,1), nrow=3, ncol=3, byrow=TRUE)
A
```

```
##      [,1] [,2] [,3]
## [1,]    2    3    7
## [2,]    1    6    2
## [3,]    3    5    1
```

```
B = matrix(c(3,2,9,0,7,8,5,8,2), nrow=3, ncol=3, byrow=TRUE)
B
```

```
##      [,1] [,2] [,3]
## [1,]    3    2    9
## [2,]    0    7    8
## [3,]    5    8    2
```

```
#Que:2 Find the matrix product.
#Answer:
C = A%*%B
C
```

```
##      [,1] [,2] [,3]
## [1,]   41   81   56
## [2,]   13   60   61
## [3,]   14   49   69
```

```
#Que:3What does correlation coefficient represent? How do you find the
correlation between (i) two categorical variables and (ii)categorical
variable and continuous variable.
#Answer:
#The correlation coefficient of two variables in a data set equals to their
covariance divided by the product of their individual standard deviations. It
is a normalized measurement of how the two are linearly related.

#(i)Frequency table, relative frequency table and bar chart etc.

#(ii)Z-test/t-test (Either test assess whether mean of two groups are
statistically different from each other or not
```

```r
#Que4:Load the Animals data set from the "MASS" package
#Answer:
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.2
```

```r
data(Animals)
head(Animals)
```

```
##                      body brain
## Mountain beaver      1.35   8.1
## Cow                465.00 423.0
## Grey wolf           36.33 119.5
## Goat                27.66 115.0
## Guinea pig           1.04   5.5
## Dipliodocus      11700.00  50.0
```

```r
#Que:5 Find the correlation coefficient of brain and body in this data set
and comment on the relation between them
#Answer:
body <- Animals$body
brain <- Animals$brain
cor(brain, body)
```

```
## [1] -0.005341163
```

```r
cor(Animals)
```

```
##                body        brain
## body   1.000000000 -0.005341163
## brain -0.005341163  1.000000000
```

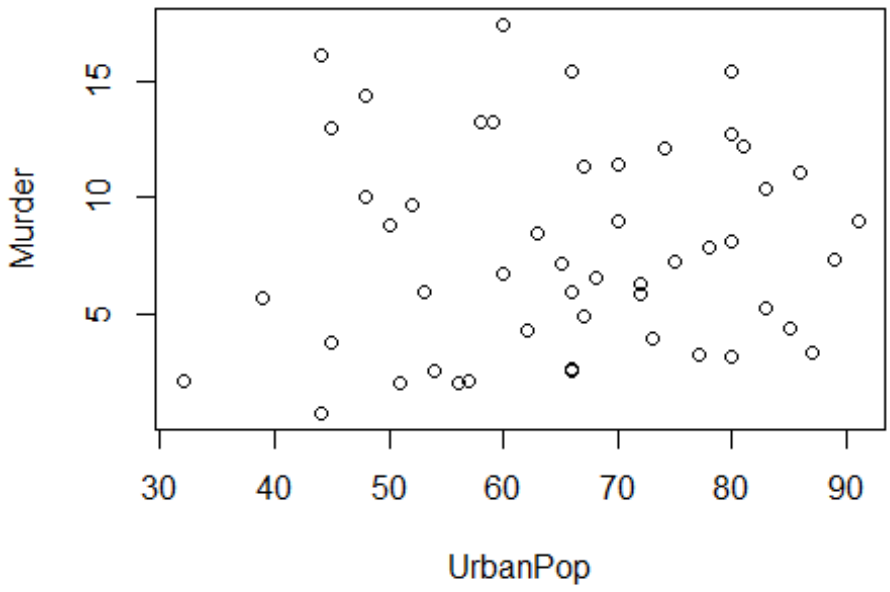
```r
#Que:6 Load USArrests data set. Comment on the distribution of the variables.
Is there any relation between UrbanPop and three crimes
#Answer:
summary(USArrests)
```

```
##      Murder          Assault         UrbanPop          Rape
##  Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
##  1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
##  Median : 7.250   Median :159.0   Median :66.00   Median :20.10
##  Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
##  3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
##  Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
```

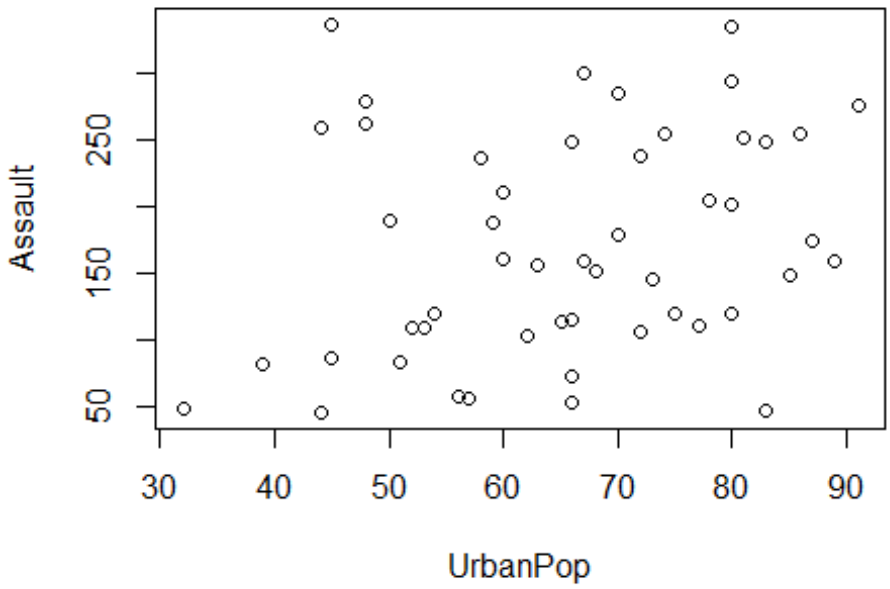
```r
cor(USArrests)
```

```
##              Murder   Assault   UrbanPop      Rape
## Murder   1.00000000 0.8018733 0.06957262 0.5635788
## Assault  0.80187331 1.0000000 0.25887170 0.6652412
## UrbanPop 0.06957262 0.2588717 1.00000000 0.4113412
## Rape     0.56357883 0.6652412 0.41134124 1.0000000
```

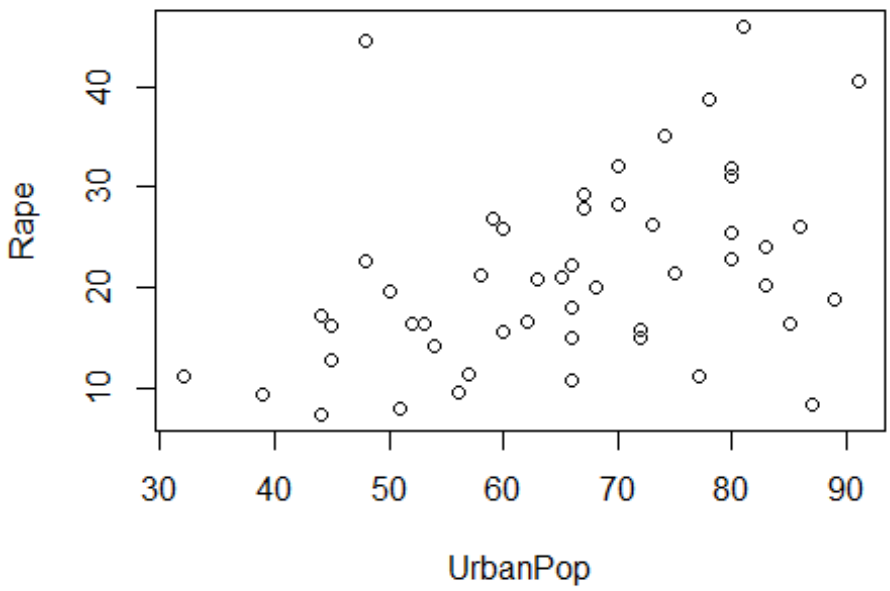
```r
#Running plot for these combinations, Murder and Assault do not appear to
have a relation to UrbanPop. The distribution of plot points are scattered to
the point that they do not appear to correlate to UrbanPop.
with(USArrests, plot(UrbanPop, Murder))
```




```r
with(USArrests, plot(UrbanPop, Assault))
```




```
#Plotting Rape against UrbanPop - there are some outliers, but there does
seem to be a relationship between Rape and UrbanPop. There is an indication
of a correlation and more analysis could be performed
with(USArrests, plot(UrbanPop, Rape))
```




```
#Que:7 Which states has most and least assault, murder, and rape arrests.
#Answer:
# most and least assault
x <- which(USArrests$Assault == max(USArrests$Assault))
rownames(USArrests)[x]

## [1] "North Carolina"

x <- which(USArrests$Assault == min(USArrests$Assault))
rownames(USArrests)[x]

## [1] "North Dakota"

# most and least murder
x<-which(USArrests$Murder == max(USArrests$Murder))
rownames(USArrests)[x]

## [1] "Georgia"

x <- which(USArrests$Murder == min(USArrests$Murder))
rownames(USArrests)[x]

## [1] "North Dakota"

# most and least rape
x<-which(USArrests$Rape == max(USArrests$Rape))
rownames(USArrests)[x]

## [1] "Nevada"
```

```r
x <- which(USArrests$Rape == min(USArrests$Rape))
rownames(USArrests)[x]
```

```
## [1] "North Dakota"
```

```r
#Que:8 List the states which have assault arrests more than median of the
country.
#Answer:
assault.median = median(USArrests$Assault)
assault.median
```

```
## [1] 159
```

```r
subset(USArrests, Assault > assault.median, select= c(UrbanPop, Assault))
```

```
##                UrbanPop Assault
## Alabama              58     236
## Alaska               48     263
## Arizona              80     294
## Arkansas             50     190
## California           91     276
## Colorado             78     204
## Delaware             72     238
## Florida              80     335
## Georgia              60     211
## Illinois             83     249
## Louisiana            66     249
## Maryland             67     300
## Michigan             74     255
## Mississippi          44     259
## Missouri             70     178
## Nevada               81     252
## New Mexico           70     285
## New York             86     254
## North Carolina       45     337
## Rhode Island         87     174
## South Carolina       48     279
## Tennessee            59     188
## Texas                80     201
## Wyoming              60     161
```

```r
#Que:9 Which states are in the bottom 25% of murder
#Answer:
bottomQuartileMurderRate <- quantile(USArrests$Murder)[2]
bottomQuartileMurderRate
```

```
##   25%
## 4.075
```

```r
subset(USArrests, Murder < bottomQuartileMurderRate, select= c(UrbanPop,
Murder))
```

```
##                UrbanPop Murder
## Connecticut          77    3.3
## Idaho                54    2.6
## Iowa                 57    2.2
## Maine                51    2.1
## Minnesota            66    2.7
## New Hampshire        56    2.1
## North Dakota         44    0.8
## Rhode Island         87    3.4
## South Dakota         45    3.8
## Utah                 80    3.2
## Vermont              32    2.2
## Washington           73    4.0
## Wisconsin            66    2.6
```

```
#Que:10 Which states are in the top 25% of the murder.
#Answer:
topQuartileMurderRate <- quantile(USArrests$Murder)[4]
topQuartileMurderRate
```

```
##   75%
## 11.25
```

```
subset(USArrests, Murder > topQuartileMurderRate, select= c(UrbanPop,
Murder))
```

```
##                UrbanPop Murder
## Alabama              58   13.2
## Florida              80   15.4
## Georgia              60   17.4
## Louisiana            66   15.4
## Maryland             67   11.3
## Michigan             74   12.1
## Mississippi          44   16.1
## Nevada               81   12.2
## New Mexico           70   11.4
## North Carolina       45   13.0
## South Carolina       48   14.4
## Tennessee            59   13.2
## Texas                80   12.7
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this: