

# Sudeep\_32

Sudeep Kharel

2024-05-31

## Q.No.6

### Set seed for reproducibility

### Generate random data for 200 cases

a.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##   filter, lag

## The following objects are masked from 'package:base':
##   intersect, setdiff, setequal, union

library(ggplot2)

set.seed(32)

data <- data.frame(
  age = sample(10:99, 200, replace = TRUE),
  sex = sample(c("male", "female"), 200, replace = TRUE),
  education = sample(c("No education", "Primary", "Secondary", "Beyond Secondary"), 200, replace = TRUE),
  socio_economic_status = sample(c("Low", "Middle", "High"), 200, replace = TRUE),
  bmi = runif(200, 14, 38)
)
```

## Q.No.7

This table is showing the Mean and Standard Deviation of the speed of wind everyday from

month May to September (represented by 5 to 9) in the year 1973.

```
aq <- airquality
wind_by_month <- sapply(split(aq, aq$Month), FUN = function(x) c(mean(x$Wind), sd(x$Wind)))
wind_by_month_df <- as.data.frame(t(wind_by_month))
colnames(wind_by_month_df) <- c('Mean', 'SD')
print(wind_by_month_df)

##           Mean           SD
## 5 11.622581 3.531450
## 6 10.266667 3.769234
## 7  8.941935 3.035981
## 8  8.793548 3.225930
## 9 10.180000 3.461254
```

a.

Using Shapiro-Wilk test to perform goodness-of-fit test on wind variable to check if it

follows normal distribution or not.]

```
library(stats)
hist(aq$Wind)

Histogram of aq$Wind

Frequency
30
20
10
0
0 5 10 15 20
aq$Wind

shapiro.test(aq$Wind)

##
## Shapiro-Wilk normality test
##
## data: aq$Wind
## W = 0.98575, p-value = 0.1178
```

Since the W value is closer to 1 and p-value is >0.05, I conclude that the variable Wind is normally distributed.

b.

Goodness-of-fit test to check if variances among different groups(months) are equal or not.

```
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##   recode

aq$Month <- factor(aq$Month)
leveneTest(aq$Wind ~ aq$Month, data = aq, center = mean)

      Df      F value      Pr(>F)
group  4      0.1859413 0.9454222
      148             NA             NA
2 rows
```

Since p-value > 0.05, we accept the null hypothesis that means equal variance among months.

c.

Normality test of each month.

```
with(aq, shapiro.test(Wind[Month == 5]))

##
## Shapiro-Wilk normality test
##
## data: Wind[Month == 5]
## W = 0.968, p-value = 0.4659

with(aq, shapiro.test(Wind[Month == 6]))

##
## Shapiro-Wilk normality test
##
## data: Wind[Month == 6]
## W = 0.96858, p-value = 0.501

with(aq, shapiro.test(Wind[Month == 7]))

##
## Shapiro-Wilk normality test
##
## data: Wind[Month == 7]
## W = 0.95003, p-value = 0.1564

with(aq, shapiro.test(Wind[Month == 8]))

##
## Shapiro-Wilk normality test
##
## data: Wind[Month == 8]
## W = 0.98533, p-value = 0.937

with(aq, shapiro.test(Wind[Month == 9]))

##
## Shapiro-Wilk normality test
##
## data: Wind[Month == 9]
## W = 0.97853, p-value = 0.7852
```

Since the W value is closer to 1 and p-value is >0.05 in each month, I conclude that variable Wind in each Month is normally distributed. So classical one-way anova must be used

## Q.No.10

```
dataset <- iris
```

10.a),b),c)

Putting the iris data set and creating a hierarchical clustering using these methods.

```
hcasingle <- hclust(dist(dataset),method = "single")

## Warning in dist(dataset): NAs introduced by coercion

hcacomplete <- hclust(dist(dataset),method = "complete")

## Warning in dist(dataset): NAs introduced by coercion

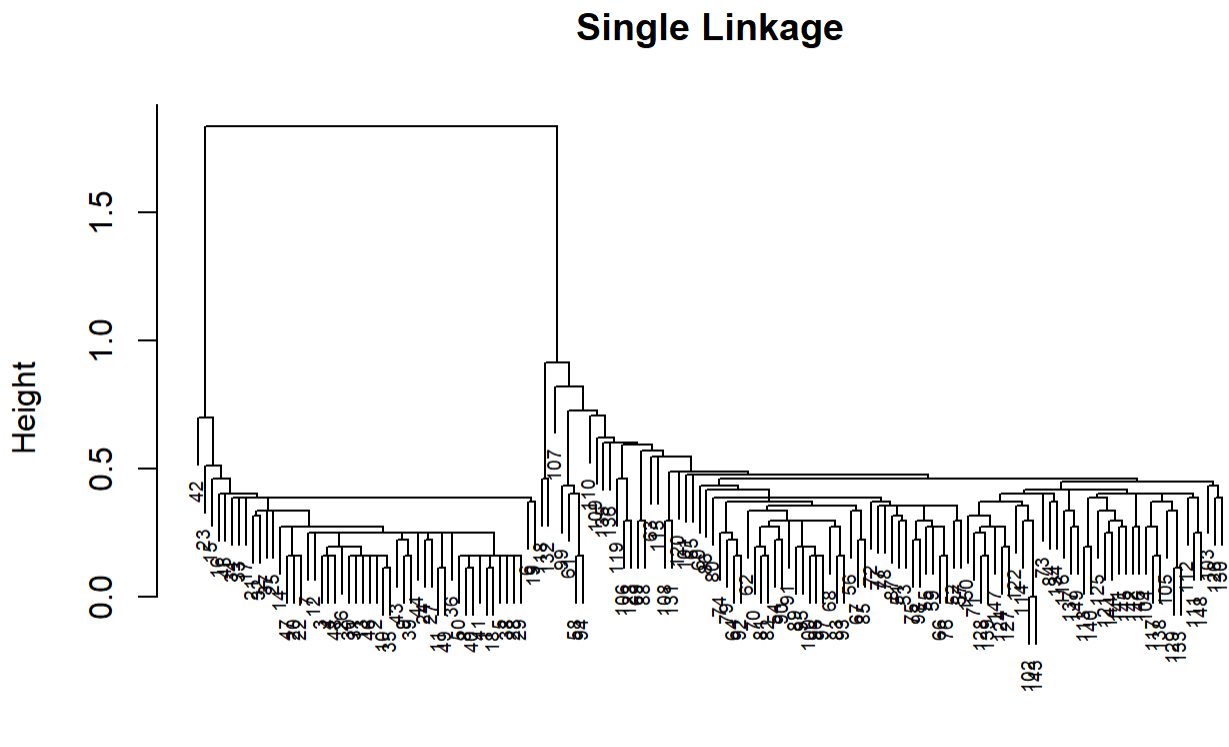
hcaaverage <- hclust(dist(dataset),method = "average")

## Warning in dist(dataset): NAs introduced by coercion
```

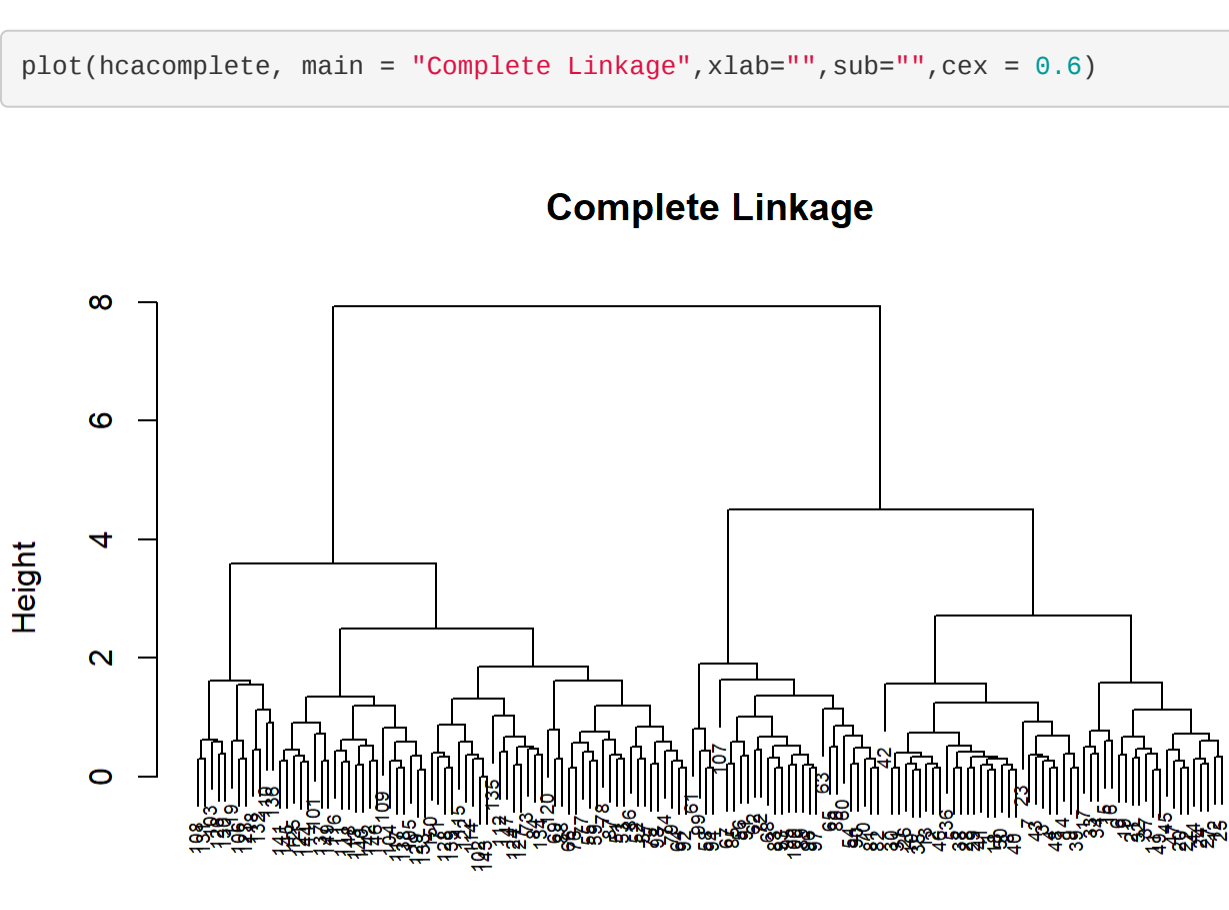
Getting the dendrogram using three different models i.e. single linkage, complete linkage,

average linkage

```
plot(hcasingle, main = "Single Linkage",xlab="",sub="",cex = 0.6)
```



```
plot(hcacomplete, main = "Complete Linkage",xlab="",sub="",cex = 0.6)
```



```
plot(hcaaverage, main = "Average Linkage",xlab="",sub="",cex = 0.6)
```

