cars

HImanshu Pareek

August 27, 2017

## Computing means for given data

Given data is for distances covered by various cars with various speeds.

reading data into R filename is "CARS.csv"

filename <- "CARS.csv"  
cars <- read.csv(filename, row.names = 1, na.strings = c("Not Available", "NOT AVAILABLE", "NULL", "NA", "na", "not available", "-", "\_", "X"))  
  
head (cars)

## speed dist  
## 1 4 2  
## 2 4 10  
## 3 7 4  
## 4 7 22  
## 5 8 16  
## 6 9 10

Removing na values (if any)

cars <- cars[complete.cases(cars), ]  
head (cars)

## speed dist  
## 1 4 2  
## 2 4 10  
## 3 7 4  
## 4 7 22  
## 5 8 16  
## 6 9 10

We will group data with respect to speed of the car

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.3.3

##   
## 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

cars <- group\_by (cars,speed)  
head (cars)

## # A tibble: 6 x 2  
## # Groups: speed [4]  
## speed dist  
## <int> <int>  
## 1 4 2  
## 2 4 10  
## 3 7 4  
## 4 7 22  
## 5 8 16  
## 6 9 10

We will thtan take average of speeds in each group

cars <- summarize(cars, mean\_dist = mean(dist))  
head (cars)

## # A tibble: 6 x 2  
## speed mean\_dist  
## <int> <dbl>  
## 1 4 6.0  
## 2 7 13.0  
## 3 8 16.0  
## 4 9 10.0  
## 5 10 26.0  
## 6 11 22.5

Now we will calculate weighted means (AM, GM and HM)

weighted\_am <- weighted.mean(cars$mean\_dist, cars$speed)  
weighted\_gm <- exp(weighted.mean(log(cars$mean\_dist), cars$speed))  
weighted\_hm <- 1 / (weighted.mean(1 / (cars$mean\_dist), cars$speed))

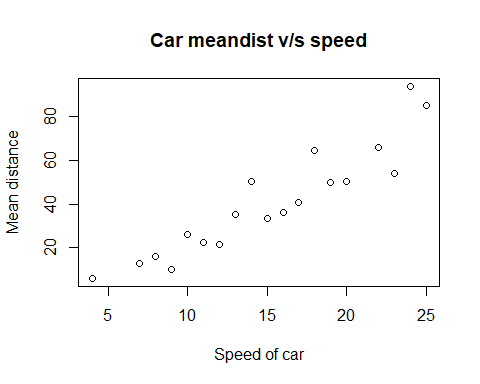
printing weighted am, gm and hm

sprintf (  
 "%s = %f",  
 c("WEIGHTED AM", "WEIGHTED GM", "WEIGHTED HM"),  
 c(weighted\_am, weighted\_gm, weighted\_hm)  
)

## [1] "WEIGHTED AM = 49.532520" "WEIGHTED GM = 42.693949"  
## [3] "WEIGHTED HM = 34.317100"

Here range and observations both are different, so we will consider GM. We can further visualize data using plot

plot (  
 cars$speed,  
 cars$mean\_dist,  
 main = "Car meandist v/s speed",  
 xlab = "Speed of car",  
 ylab = "Mean distance"  
)



We can see that relation between mean distance and speed is approximately linear