# Frequency Distribution

* **Constructing the Frequency Distribution**
  + **First Method**

mydata<- scan()

8 62 73 85 42 68 54 38 27 32 63 68 69 75 59 52 58 36 85 88 72 52 52 63 68 29 73 29 76 29 57 46 43 28 32 9 66 72 68 42 76 38 38 39 28 19 12 78 72 92 82 72 33 92 69 28 39 85 59 68 52 85 59 76 80 72 74 54 48 29 36 10 82 58 88 68 58 46 37 29 35

mydata #Print data

Another way of entering data:

data<- c (8, 62, 73, 85, 42, 68, 54, 38 ,27, 32, 63, 68, 69, 75, 59, 52, 58,36, 85, 88, 72, 52, 52, 63, 68, 29, 73, 29, 76,29, 57, 46, 43, 28, 32, 9, 66, 72, 68, 42, 76, 38, 38, 39, 28, 19, 12, 78, 72, 92, 82, 72, 33, 92, 69, 28,39 ,85, 59, 68, 52, 85, 59, 76, 80, 72, 74, 54, 48, 29, 36, 10, 82, 58, 88, 68, 58, 46, 37, 29, 35)

data #Print data

mini<-min(data) #Min value

maxi<-max(data) #Max value

n<-length(data) #Length of data

Range<-maxi-mini #Calculating range

Range/6

k<-6 # No of classes

h<-15 # width of class interval

i<-1:6

x<-c(12+h\*(i-1)) # 12 is the midpoint of first class namely 5-19

x # print x

* + - **Calculating Frequencies of the Classes:**

f<-1:6

for(i in 1:6){

f[i]<-length(data[(x[i]-h/2) <= data & data <= (x[i]+h/2)])

}

fr.dist<-data.frame(midvalue=x,frequency=f) #Converting data in the form of table

fr.dist #Printing freq dist

* + **Second Method:**

data<c(8,9,8,6,6,8,29,13,45,67,32,15,46,78,16,54,36,25,34,64,27,65,45,38,77,55,66,43,44,30,31,75,62,15,45,49,39,28,17,40,47,54,81,84,45,46,47,38,59,50)

L<-length(data)

B<-max(data)

S<-min(data)

R<-(B-S)

K=1+3.322\*log10(50)

H<-R/K

breaks<-round(seq(6,B+H,H),0)

breaks

Classes<-cut(data,breaks,right=F);

freq.dist=table(Classes)

f.d1=cbind(freq.dist)

f.d1

## Cummulative Frequency:

cum.freq.dist=cumsum(f.d1)

c.f.d<-cbind(freq.dist,cum.freq.dist)

c.f.d

cf<-sum(freq.dist)

## Relative Frequency:

rela.freq.dist=freq.dist/cf

r.f.d=cbind(freq.dist,cum.freq.dist,rela.freq.dist)

r.f.d

## Midpoints:

i<-1:7

MidPoints<-round(c(breaks[i]+(H/2)),0)

m.p<-cbind(freq.dist,cum.freq.dist,rela.freq.dist,MidPoints)

m.p

# Histogram

h1<-hist(data, xlab="Observations", ylab="Frequency", main="",col="green") #Simple Method

Histogram<-hist(data, main="Histogram", xlab="Class Boundaries",ylab="Frequency", col="skyblue", xlim=c(0,100))

Histogram<lines(c(5,Histogram$mids,95),c(0,Histogram$counts,0),lwd=1.5,pch=18,type="o",xlim=c(0,100))

Histogram<-hist(data, main="Histogram", xlab="Class Boundaries",ylab="Frequency", col="skyblue", xlim=c(0,100), freq=F)

## For Density Curve:

par(new=T)

#For Density curve

Histogram<-hist(data,main="Histogram", xlab="Class Boundaries",prob=T,xlim=c(-10,100))

lines(density(data,adjust=0.5),xlim=c(-10,100),col="red")

# Frequency Polygon

**#superimpose a frequency polygon on top of a histogram**

Right way after modify the above graph.

data<-c(8,9,8,6,6,8,29,13,45,67,32,15,46,78,16, 54,36,25,34,64,27,65,45,38,77,55,66,43,44,30,31,

75,62,15,45,49,39,28,17,40,47,54,81,84,45,46,47,38,59,50)

breaks<-seq(-5,102,20) #plot range(x-axis)

mids<-seq(12,102,15) # mids point

h1<-hist(data, xlab="Observations", ylab="Frequency",main="",col="grey",xlim=c(-5,104))

h2<-lines(c(min(breaks),h1$mids,max(breaks)),c(0,h1$counts,0),lwd=2,type="o")

**#Another way**

# declare your variables

data <- c(10, 7, 8, 4, 5, 6, 6, 9, 5, 6, 3, 8,4, 6, 10, 5, 9, 7, 6, 2, 6, 5, 4, 8, 7, 5, 6)

# find the range

range(data)

# establish a class width

class\_width = seq(1, 11, by=2)

class\_width

# create a frequency table

data.cut = cut(data, class\_width, right=FALSE)

data.freq = table(data.cut)

cbind(data.freq)

# histogram of this data

hist(data,

breaks=class\_width,

col="slategray3",

border = "dodgerblue4",

right=FALSE,

xlab = "Scores",

main = "Histogram of Quiz Data")

# this is key to the overlay

par(new=TRUE)

# create a frequency polygon for the birth weight data

plot(data.freq, type="b")