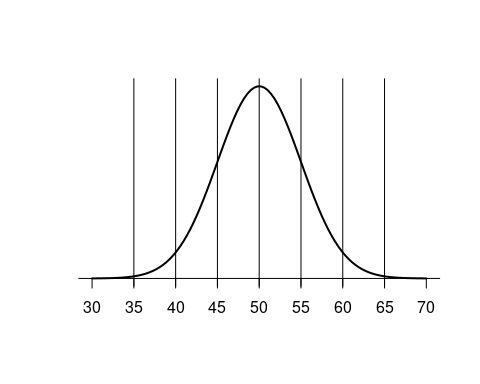
Assignment 5

Dipesh Poudel

12/8/2021

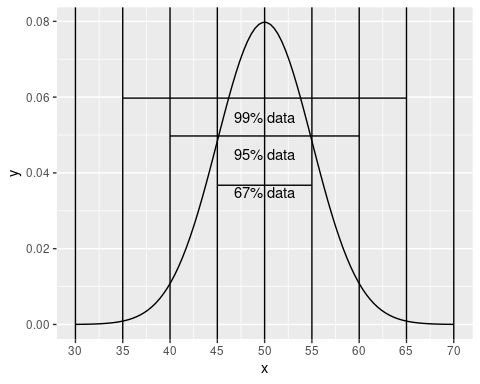
# Assignment 5

#Define mean and SD  
pop\_mean<-50  
pop\_sd<-5  
#Define lower and upper limits  
LL <-pop\_mean-pop\_sd  
UL <-pop\_mean+ pop\_sd  
#Create a sequence of 100 x values based on pop mean and sd  
x <-seq(-4,4, length=100)\*pop\_sd+pop\_mean  
y <-dnorm(x, pop\_mean, pop\_sd)  
plot(x,y, type="l", lwd=2, axes=F, xlab="", ylab="")  
sd\_axis\_bounds= 5  
axis\_bounds<-seq(-sd\_axis\_bounds\*pop\_sd + pop\_mean, sd\_axis\_bounds\*pop\_sd+ pop\_mean, by=pop\_sd)  
axis(side=1, at=axis\_bounds, pos=0)  
abline(v=50)  
abline(v=50+pop\_sd)  
abline(v=50-pop\_sd)  
abline(v=50+2\*pop\_sd)  
abline(v=50-2\*pop\_sd)  
abline(v=50+3\*pop\_sd)  
abline(v=50-3\*pop\_sd)

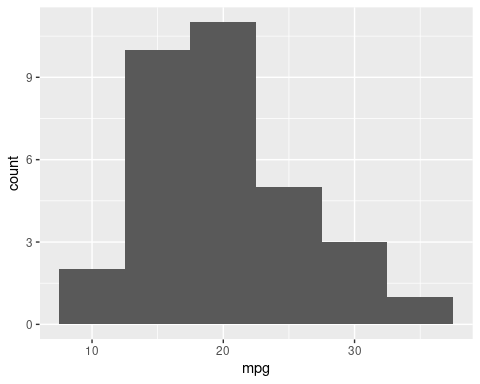


library(ggplot2)  
ggplot(data=as.data.frame(x,y),aes(x=x,y=y))+geom\_line()+scale\_x\_continuous(breaks = axis\_bounds)+geom\_vline(xintercept = seq(30,70,by=pop\_sd))+annotate(geom = "text",x = 50,y=mean(y)+0.01, label = "67% data")+annotate(geom = "segment",x = 45,y=mean(y)+0.012,xend = 55,yend = mean(y)+0.012)+annotate(geom = "text",x = 50,y=mean(y)+0.02, label = "95% data")+annotate(geom = "segment",x = 40,y=mean(y)+0.025,xend = 60,yend = mean(y)+0.025)+annotate(geom = "text",x = 50,y=mean(y)+0.03, label = "99% data")+annotate(geom = "segment",x = 35,y=mean(y)+0.035,xend = 65,yend = mean(y)+0.035)

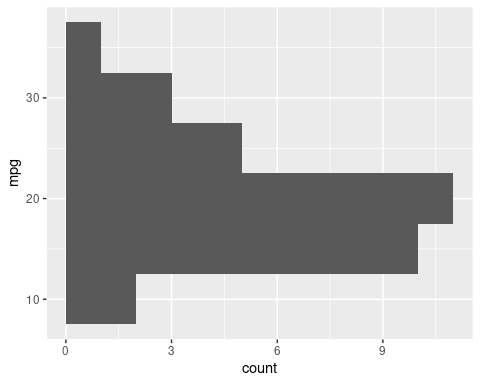
## Warning in as.data.frame.numeric(x, y): 'row.names' is not a character vector of  
## length 100 -- omitting it. Will be an error!



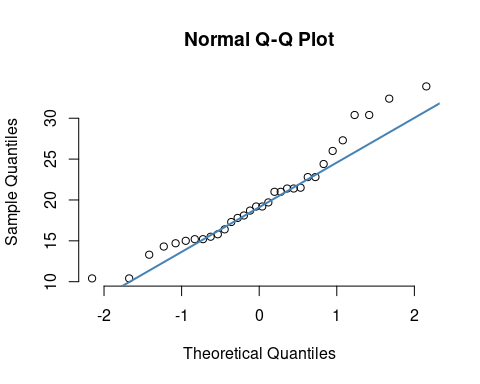
cars\_data<-mtcars  
ggplot(data=cars\_data,aes(x=mpg))+geom\_histogram(binwidth = 5)



ggplot(data = cars\_data,aes(mpg))+geom\_histogram(binwidth = 5)+coord\_flip()



cars\_data<-mtcars  
qqnorm(cars\_data$mpg, pch = 1, frame = FALSE)  
qqline(cars\_data$mpg, col = "steelblue", lwd = 2)



shapiro.test(cars\_data$mpg)

##   
## Shapiro-Wilk normality test  
##   
## data: cars\_data$mpg  
## W = 0.94756, p-value = 0.1229

Null Hypothesis(H0): The data follows normal distribution Alternative Hypothesis(H1): The data does not follow normal distribution Since the p-value obtained from the test is not less than 0.05 we can say that the data comes from normal distribution i.e Null Hypothesis is accepted.

Since the number of samples (32)<50 we can not use Kolmogorov-Smirnov. Thus we have to use Shapiro-Wilk normality test.