Sampling

Dr. B

Saturday, November 15, 2014

## Setup

##Load some packages  
if (!require("Hmisc")){  
 install.packages("Hmisc")  
}  
  
if (!require("xtable")){  
 install.packages("table")  
}  
  
##Clear the environment  
rm(list=ls())  
  
#Set variables to use  
set.seed(2345) ##Set the seed for reproducibility  
nbr = 100 ##Number of observations for each group (3 groups)  
sampsize = nbr #Size of samples  
nbrsamples = 10000 #how many samples  
withreplace = 1 #With replacement 0 or 1 (1 for bootstrap)   
conf = .95 ##COnfidence Interval

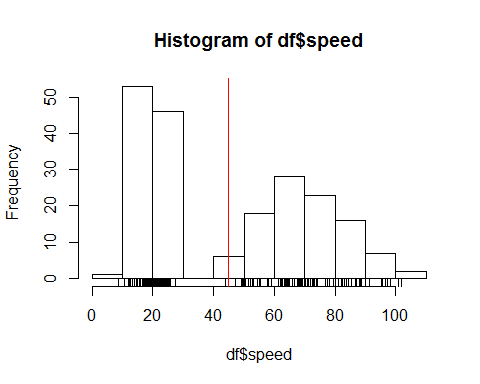
## Generate random data

##Create the data attributes  
##Speed  
s1<- rnorm(nbr,70,15)  
s2<- rep(NA,nbr)  
s3<- rnorm(nbr,20,4)  
speed<-c(s1,s2,s3)  
rm(s1,s2,s3)  
  
##Cycles  
c1<- rbinom(nbr,3,.6)  
c2<- rbinom(nbr,3,.4)  
c3<- rep(NA,nbr)  
cycle<-c(c1,c2,c3)  
rm(c1,c2,c3)  
  
##Wear  
w1<- rep(NA,nbr)  
w2<- rchisq(nbr,4)  
w3<- rchisq(nbr,7)  
wear<-c(w1,w2,w3)  
rm(w1,w2,w3)  
  
##Establish Group Numbers  
group<- c(rep(1,nbr),rep(2,nbr),rep(3,nbr))  
  
##Make a dataframe  
df <-data.frame(speed=speed, cycle=cycle, wear=wear, group=group)  
rm(speed,cycle,wear,group)

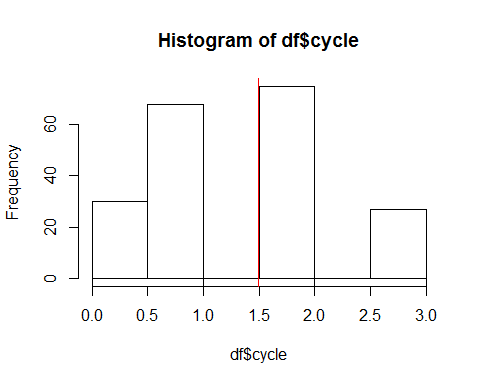
### Histograms

Next we make a histogram for each variable:

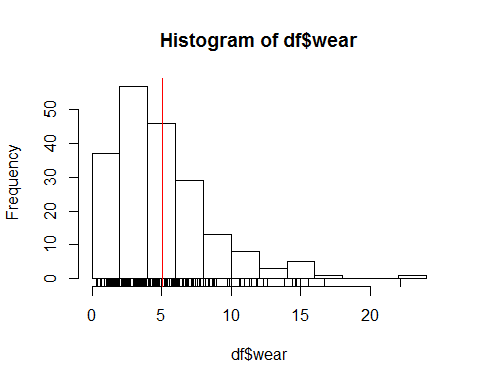
hist(df$speed)  
rug(df$speed)  
abline(v=mean(df$speed,na.rm=1),col="red")



hist(df$cycle)  
rug(df$cycle)  
abline(v=mean(df$cycle,na.rm=1),col="red")



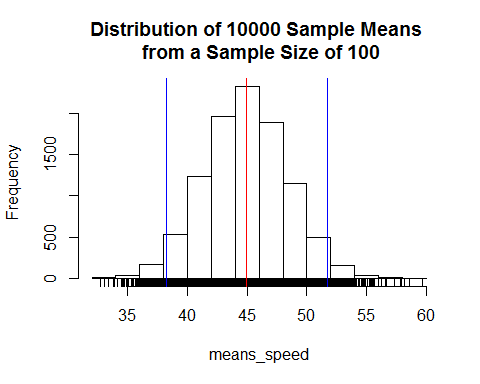
hist(df$wear)  
rug(df$wear)  
abline(v=mean(df$wear,na.rm=1),col="red")



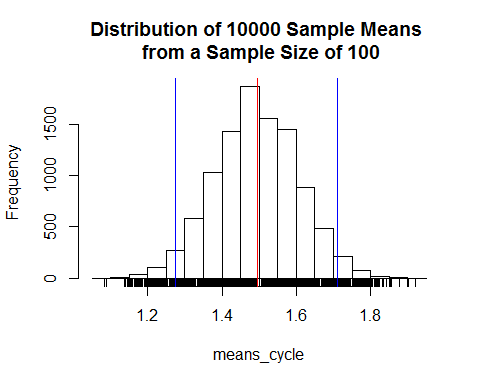
### Sampling

Next we will take 10^{4} samples of size 100 of each variable and make a histogram of each

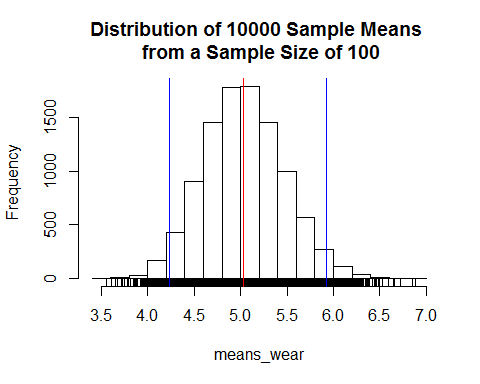
maintitle=paste0("Distribution of ",nbrsamples," Sample Means \n from a Sample Size of ", sampsize)  
  
##One way to Sample  
#means\_speed <- replicate(nbrsamples,mean(sample(df$speed,sampsize,replace=withreplace),na.rm=1))  
#hist(means\_speed, main=maintitle)  
#rug(means\_speed)  
#abline(v=mean(df$speed,na.rm=1),col="red")  
##quantile(medians, c(.025,.975))  
#abline(v=quantile(means\_speed, c((1 - conf)/2)),col="blue")  
#abline(v=quantile(means\_speed, c((1 - conf)/2 +conf)),col="blue")  
  
##Bootstraping   
resamples <-matrix(sample(df$speed,sampsize\*nbrsamples,replace=withreplace),nbrsamples,sampsize)  
means\_speed <-apply (resamples,1,mean, na.rm=1)  
hist(means\_speed, main=maintitle)  
rug(means\_speed)  
abline(v=mean(df$speed,na.rm=1),col="red")  
#quantile(medians, c(.025,.975))  
abline(v=quantile(means\_speed, c((1 - conf)/2)),col="blue")  
abline(v=quantile(means\_speed, c((1 - conf)/2 +conf)),col="blue")



##One way to Sample  
#means\_cycle <- replicate(nbrsamples,mean(sample(df$cycle,sampsize,replace=withreplace),na.rm=1))  
#hist(means\_cycle, main=maintitle)  
#rug(means\_cycle)  
#abline(v=mean(df$cycle,na.rm=1),col="red")  
##quantile(medians, c(.025,.975))  
#abline(v=quantile(means\_cycle, c((1 - conf)/2)),col="blue")  
#abline(v=quantile(means\_cycle, c((1 - conf)/2 +conf)),col="blue")  
  
##Bootstrapping  
resamplescycle <-matrix(sample(df$cycle,sampsize\*nbrsamples,replace=withreplace),nbrsamples,sampsize)  
means\_cycle <-apply (resamplescycle,1,mean, na.rm=1)  
hist(means\_cycle, main=maintitle)  
rug(means\_cycle)  
abline(v=mean(df$cycle,na.rm=1),col="red")  
#quantile(medians, c(.025,.975))  
abline(v=quantile(means\_cycle, c((1 - conf)/2)),col="blue")  
abline(v=quantile(means\_cycle, c((1 - conf)/2 +conf)),col="blue")



##One way to sample  
#means\_wear <- replicate(nbrsamples,mean(sample(df$wear,sampsize,replace=withreplace),na.rm=1))  
#hist(means\_wear, main=maintitle)  
#rug(means\_wear)  
#abline(v=mean(df$wear,na.rm=1),col="red")  
##quantile(medians, c(.025,.975))  
#abline(v=quantile(means\_cycle, c((1 - conf)/2)),col="blue")  
#abline(v=quantile(means\_cycle, c((1 - conf)/2 +conf)),col="blue")  
  
##Bootstrapping  
resampleswear <-matrix(sample(df$wear,sampsize\*nbrsamples,replace=withreplace),nbrsamples,sampsize)  
means\_wear <-apply (resampleswear,1,mean, na.rm=1)  
hist(means\_wear, main=maintitle)  
rug(means\_wear)  
abline(v=mean(df$wear,na.rm=1),col="red")  
##quantile(medians, c(.025,.975))  
abline(v=quantile(means\_wear, c((1 - conf)/2)),col="blue")  
abline(v=quantile(means\_wear, c((1 - conf)/2 +conf)),col="blue")



### Summary

##Let's look at summaries & boxplots  
summary(means\_speed)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 32.66 42.60 44.91 44.92 47.22 59.65

summary(means\_cycle)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.083 1.419 1.493 1.495 1.571 1.921

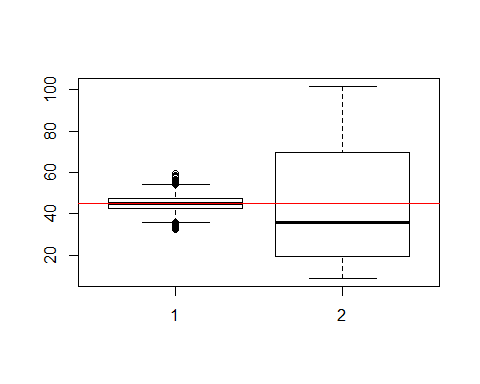
summary(means\_wear)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.554 4.735 5.024 5.038 5.322 6.883

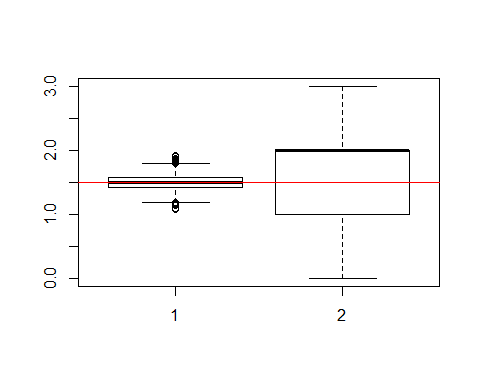
summary(df[,-4])

## speed cycle wear   
## Min. : 8.77 Min. :0.000 Min. : 0.3048   
## 1st Qu.: 19.52 1st Qu.:1.000 1st Qu.: 2.4517   
## Median : 35.73 Median :2.000 Median : 4.2859   
## Mean : 44.92 Mean :1.495 Mean : 5.0352   
## 3rd Qu.: 69.77 3rd Qu.:2.000 3rd Qu.: 6.6046   
## Max. :101.71 Max. :3.000 Max. :22.1443   
## NA's :100 NA's :100 NA's :100

boxplot(means\_speed,df$speed)  
abline(h=mean(df$speed,na.rm=1),col="red")



boxplot(means\_cycle,df$cycle)  
abline(h=mean(df$cycle,na.rm=1),col="red")



boxplot(means\_wear,df$wear)  
abline(h=mean(df$wear,na.rm=1),col="red")

