## 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)</pre>
c2<- rbinom(nbr,3,.4)</pre>
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))

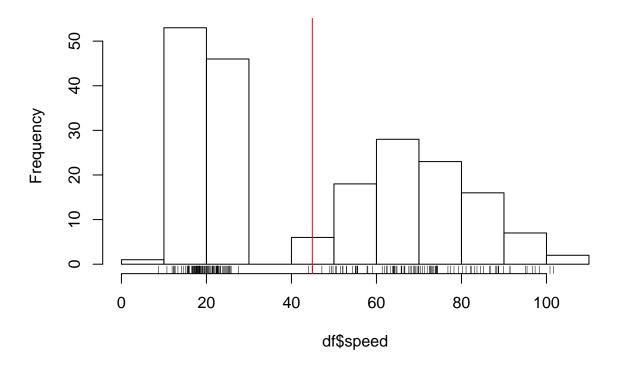
##Make a dataframe
df <-data.frame(speed=speed, cycle=cycle, wear=wear, group=group)
rm(speed,cycle,wear,group)</pre>
```

## Histograms

Next we make a histogram for each variable:

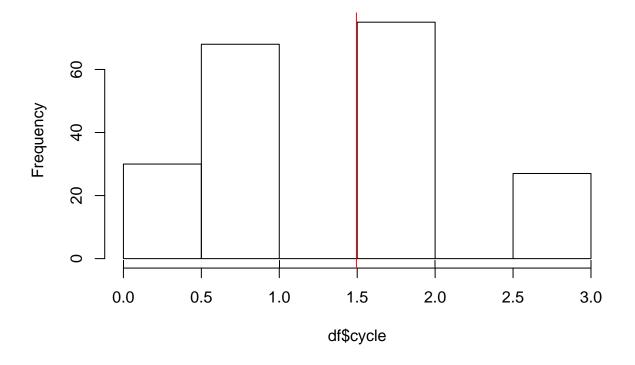
```
hist(df$speed)
rug(df$speed)
abline(v=mean(df$speed,na.rm=1),col="red")
```

## Histogram of df\$speed



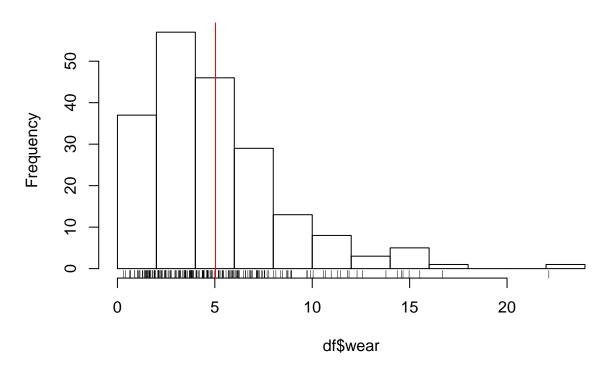
```
hist(df$cycle)
rug(df$cycle)
abline(v=mean(df$cycle,na.rm=1),col="red")
```

## Histogram of df\$cycle



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

## Histogram of df\$wear

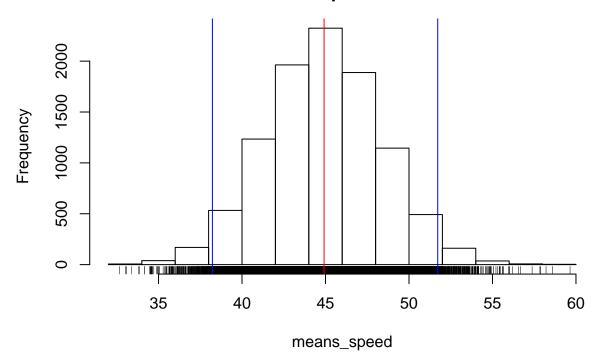


### Sampling

Next we will take 10<sup>4</sup> samples of size 100 of each variable and make a histogram of each

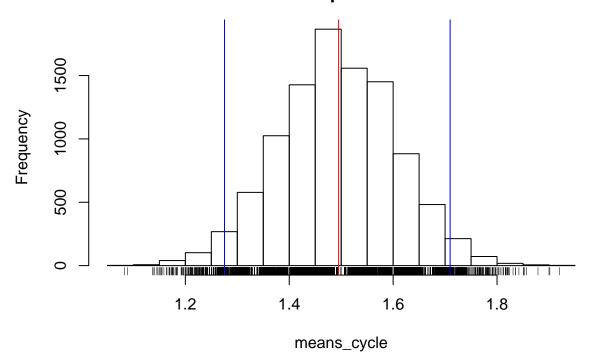
```
maintitle=paste0("Distribution of ",nbrsamples," Sample Means \n from a Sample Size of ", sampsize)
##One way to Sample
\#means\_speed \leftarrow replicate(nbrsamples, mean(sample(df\$speed, sampsize, replace=with replace), na.rm=1))
#hist(means_speed, main=maintitle)
#ruq(means_speed)
#abline(v=mean(df$speed,na.rm=1),col="red")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

# Distribution of 10000 Sample Means from a Sample Size of 100



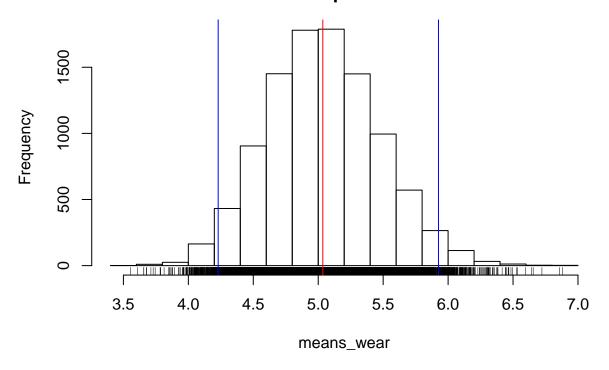
```
##One way to Sample
\#means\_cycle \leftarrow replicate(nbrsamples, mean(sample(df$cycle, sampsize, replace=with replace), na.rm=1))
#hist(means_cycle, main=maintitle)
#ruq(means_cycle)
#abline(v=mean(df$cycle,na.rm=1),col="red")
##quantile(medians, c(.025,.975))
#abline(v=quantile(means_cycle, c((1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

# Distribution of 10000 Sample Means from a Sample Size of 100



```
##One way to sample
\#means\_wear \leftarrow replicate(nbrsamples, mean(sample(df$wear, sampsize, replace=with replace), na.rm=1))
#hist(means_wear, main=maintitle)
#ruq(means_wear)
#abline(v=mean(df$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$wear,sampsize*nbrsamples,replace=withreplace),nbrsamples,sampsize)
means_wear <-apply (resampleswear,1,mean, na.rm=1)</pre>
hist(means_wear, main=maintitle)
rug(means_wear)
abline(v=mean(df$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")
```

# Distribution of 10000 Sample Means from a Sample Size of 100



## **Summary**

##

speed

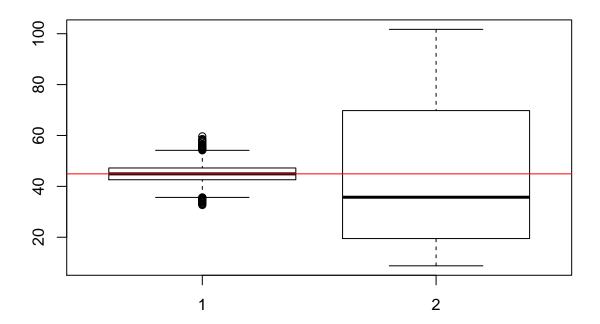
cycle

```
##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.
           1.419
##
     1.083
                    1.493
                             1.495
                                     1.571
                                             1.921
summary(means_wear)
     Min. 1st Qu. Median
                             Mean 3rd Qu.
##
                                              Max.
           4.735
                    5.024
                             5.038
                                    5.322
##
     3.554
                                            6.883
summary(df[,-4])
```

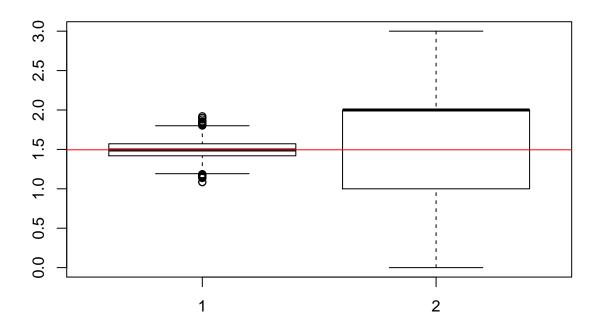
wear

```
## Min. : 8.77
                    Min.
                           :0.000
                                    Min.
                                           : 0.3048
##
   1st Qu.: 19.52
                    1st Qu.:1.000
                                    1st Qu.: 2.4517
                    Median :2.000
  Median : 35.73
                                    Median : 4.2859
##
  Mean
          : 44.92
                    Mean
                           :1.495
                                    Mean
                                          : 5.0352
   3rd Qu.: 69.77
                    3rd Qu.:2.000
                                    3rd Qu.: 6.6046
##
                           :3.000
##
  Max.
          :101.71
                    Max.
                                    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")
```

