

# assignment1

August 27, 2023

## 1 Assignment 1

### 1.1 Introduction

There are 8 groups named as soprano.1,soprano.2,alto.1,alto.2,tenor.1,tenor.2,bass.1,bass.2 which contains height in inches of the singers in the New York Choral Society in 1979. The names of groups are as components. The first 4 components are female voices and the last 4 are male voices. They are in order of decreasing pitch. Pitch is similar to frequency which describe a voice is high or low. The problem is to find out the relationship between height and gender or height and voice type.

```
[ ]: load('classdata.RData') #Read data from document
```

```
[ ]: names(singers)
```

1. 'soprano.1' 2. 'soprano.2' 3. 'alto.1' 4. 'alto.2' 5. 'tenor.1' 6. 'tenor.2' 7. 'bass.1' 8. 'bass.2'

### 1.2 Plot and analysis

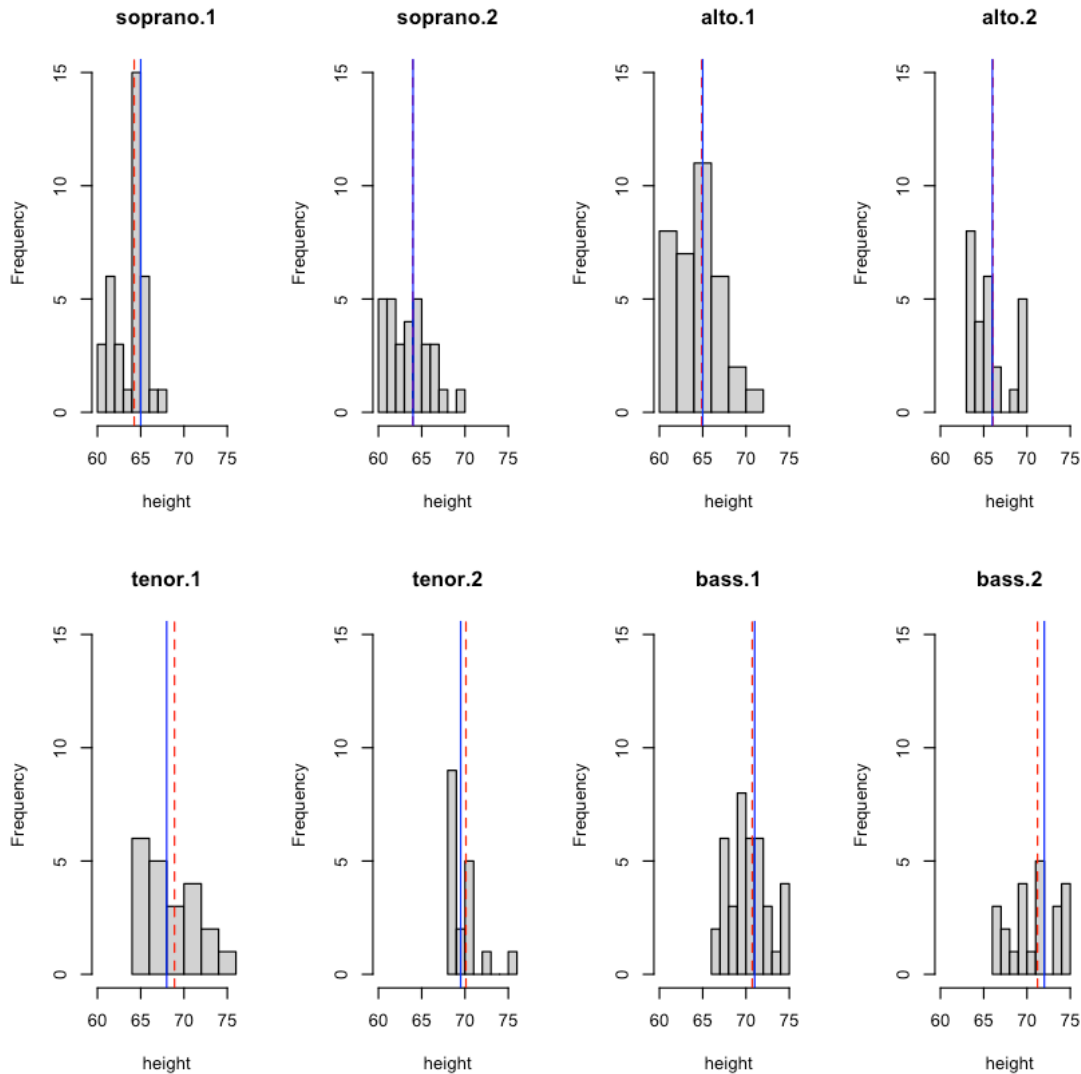
First, take a brief look at the height distributions of each groups. The plots are shown below. The medians and means of height are red lines and blue line respectively. The red and blue lines seem to move slightly to right from left graphic to right graphic in each row and the median and mean are obviously shown larger in each column.

```
[ ]: meanheight=NULL
medianheight=NULL
xname=NULL
varheight=NULL
numpe=NULL
par(mfrow=c(2,4),bg='white')
for (i in 1:length(singers)){
  height<-unlist(singers[i])
  meanheight<-c(meanheight,mean(height))
  medianheight<-c(medianheight,median(height))
  varheight<-c(varheight,var(height))
  xname<-c(xname,names(singers[i]))
  numpe<-c(numpe,length(height))
  par(bg='white')
  hist(height,main=names(singers[i]),breaks=8,xlim=c(60,76),ylim=c(0,15))
}
```

```

abline(v=mean(height),col='red',lty=2)
abline(v=median(height),col='blue')
}

```



To have a much clearer view of the median and mean of each group, the graphic below shows medians and means with blue dots and red lines. The picture shows a relationship between height and gender as well as height and voice pitch. The height of males is larger than the height of females. Meanwhile, there is also a phenomenon that while pitch is increasing, height is decreasing.

```

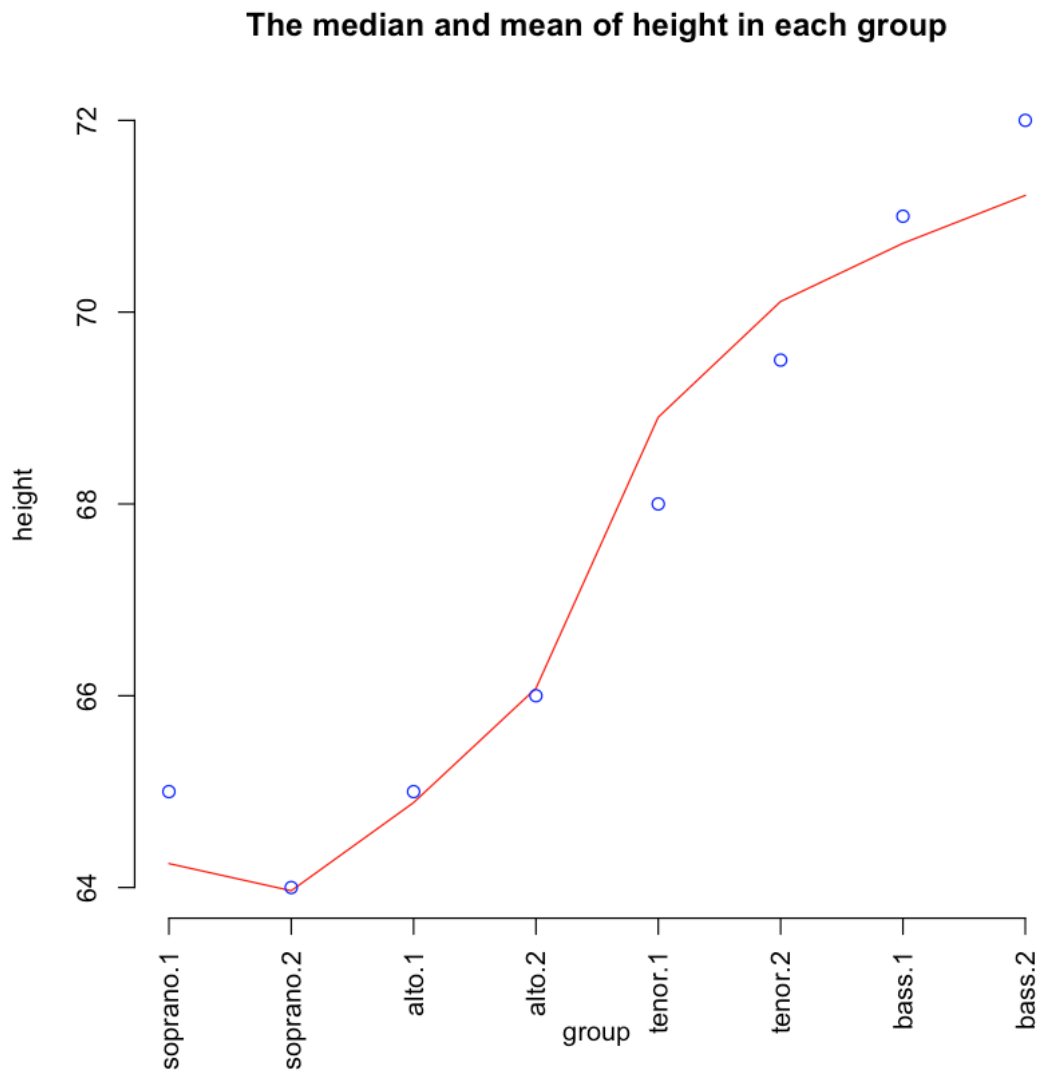
[ ]: par(mfrow=c(1,1),new=FALSE)
par(bg='white')
plot(meanheight,col='red',type='l',ylim=c(64,72),xlab='',ylab='',axes=FALSE,main='The_
median and mean of height in each group')

```

```

par(new=TRUE)
plot(medianheight,col='blue',ylim=c(64,72),xlab='group',ylab='height',axes=FALSE)
axis(2)
axis(1,at=seq_along(medianheight),labels =xname,las=2)

```



Besides, the variance of male distribution which is 12.7957718239997 larger than 12.7830128178062 seems to be larger but the difference is very slight. There is also no obvious relationship between variance of height and pitch from the graphic

```

[ ]: print(paste('female variance:',sum(varheight*(numpe-1)[1:4])/(sum(numpe[1:
  ↪4])-1)))#female
print(paste('male variance:',sum(varheight*(numpe-1)[5:8])/(sum(numpe[5:
  ↪8])-1)))#male

```

```

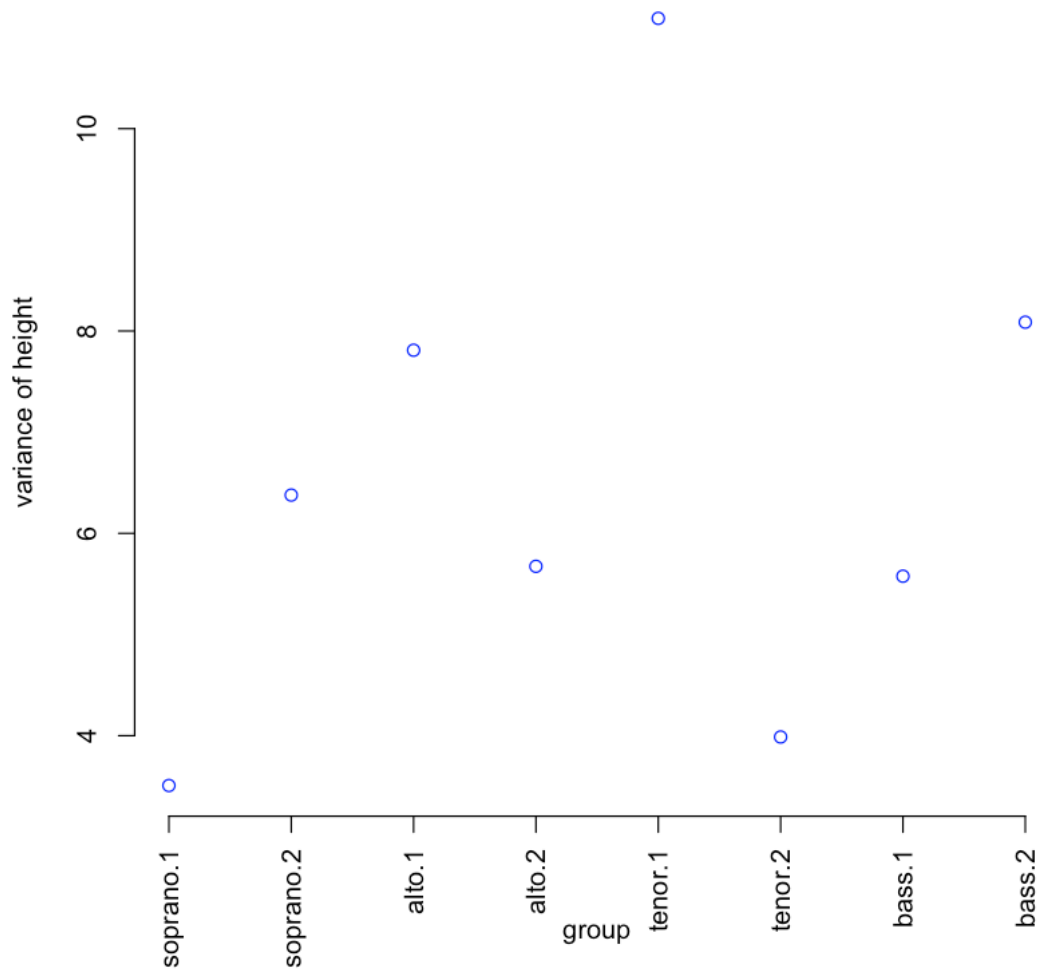
par(mfrow=c(1,1),new=FALSE)
par(bg='white')
plot(varheight,col='blue',xlab='group',ylab='variance of height',
      ylim=range(varheight),axes=FALSE,main='The variance of height in each group')
axis(2)
axis(1,at=seq_along(varheight),labels =xname,las=2)
#abline(lm(varheight~c(1:8)))
par(new=TRUE)

```

[1] "female variance: 12.7830128178062"

[1] "male variance: 12.7957718239997"

### The variance of height in each group



### **1.3 Conclusion**

Gender and pitch are factors which can describe height of singers. Males seem to be taller and singers with lower voice pitch seem to be taller.