**Advanced statistics for data science**

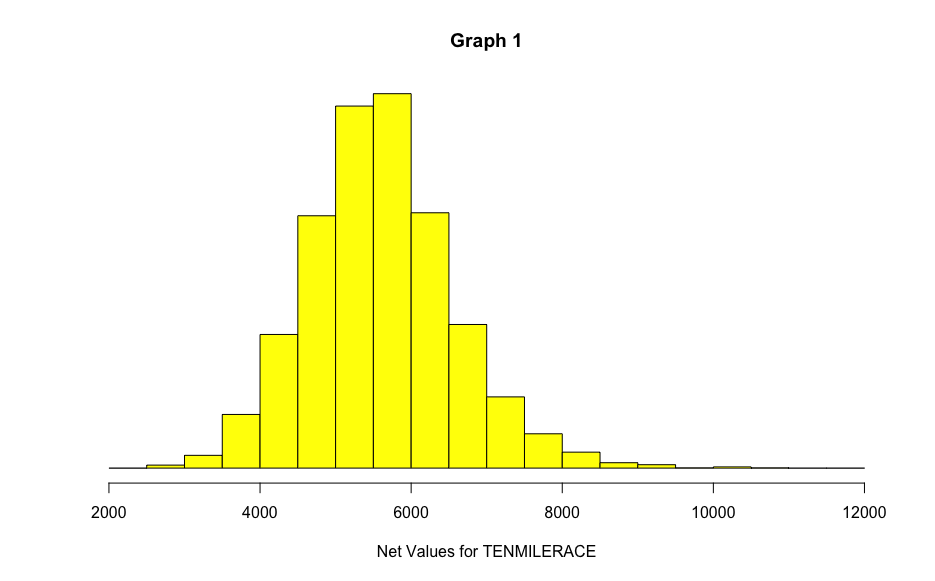
**Mini project: Histograms**

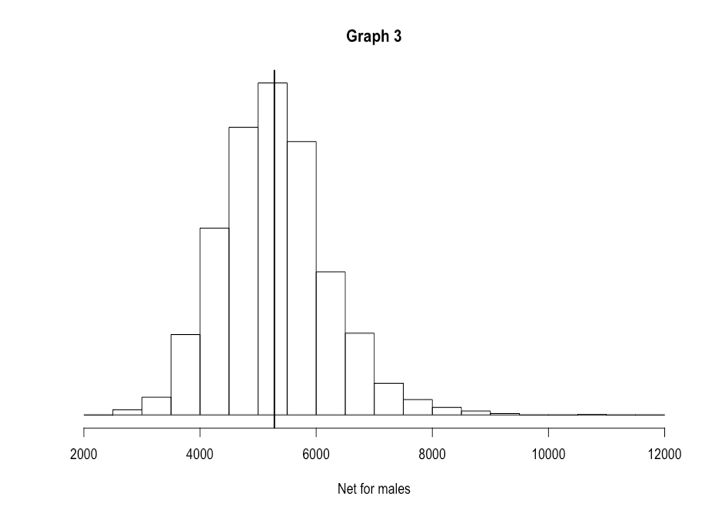
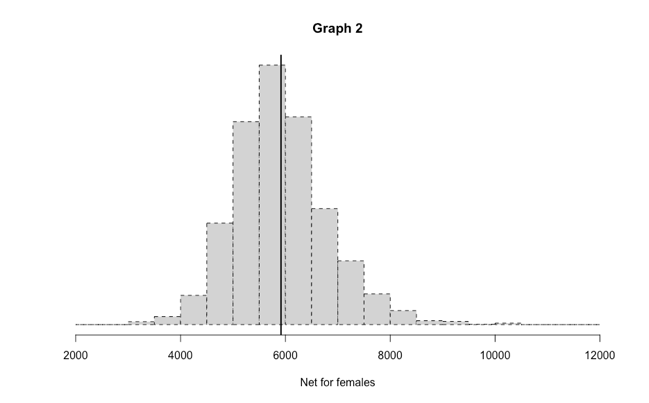
The Cherry Blossom Ten Mile Race: Net Racing Times for Women and Men

Submitted by: Submitted to:

Jayesh Haryani Dr. Tristan Whalen

JXH230004 Associate Professor of Instruction(JSOM)





a. How do the shapes of the two histograms compare? Are they the same? Can you roughly get from one to the other by shifting (adding) or scaling (stretching)?

b. How do the SDs of the two groups compare? Are they similar or different?

c. How do the averages of the two groups compare? What is the difference? How is it interpreted? Based on your answers to the previous parts, how adequate a summary would the difference between the averages be?

Ans a)

|  |  |  |  |
| --- | --- | --- | --- |
| -- | Average net | Sample variance | Standard deviation |
| Male | 5280.70 | 864866.1 | 929. 9817 |
| Female | 5916.398 | 813800.7 | 902.109 |

**R script used to perform the required instructions**

Graph 1

hist(TenMileRace$net,breaks=seq(from=2000, to=12000, by=500),

prob=TRUE,

yaxt='n',

ylab=NULL,

xlab="Net Values for TENMILERACE",

main=’Graph 1’,

col="yellow")

graph 2

> hist(TenMileRace[TenMileRace$sex=='F',]$net,breaks=seq(from=2000, to=12000, by=500),

+ prob=TRUE,

+ yaxt='n',

+ ylab=NULL,

+ xlab="Net for females",

+ main=’Graph 2’,

+ lty=2)

> abline(v=mean(TenMileRace$net[TenMileRace$sex=="F"]), lwd=2)

Graph 3

hist(TenMileRace[TenMileRace$sex=='M',]$net,breaks=seq(from=2000, to=12000, by=500),

prob=TRUE,

yaxt='n',

ylab=NULL,

xlab="Net for males",

main=’Graph 3’,

col=NULL)

abline(v=mean(TenMileRace$net[TenMileRace$sex=="M"]), lwd=2)

Average(mean) script used

mean(TenMileRace[TenMileRace$sex=='F',]$net) for females

mean(TenMileRace[TenMileRace$sex=='M',]$net) for male

Males

variance = var(TenMileRace$net[TenMileRace$sex=='M'])

Standard deviation = SD<- sqrt(var(TenMileRace$net[TenMileRace$sex=='M']))

SD

Females

variance = var(TenMileRace$net[TenMileRace$sex=='F'])

Standard deviation = SDF<-sqrt(var(TenMileRace$net[TenMileRace$sex=='F']))

SDF