

# Create data sets for Week 7

*Stat 201: Statistics I*

*March 10, 2019*

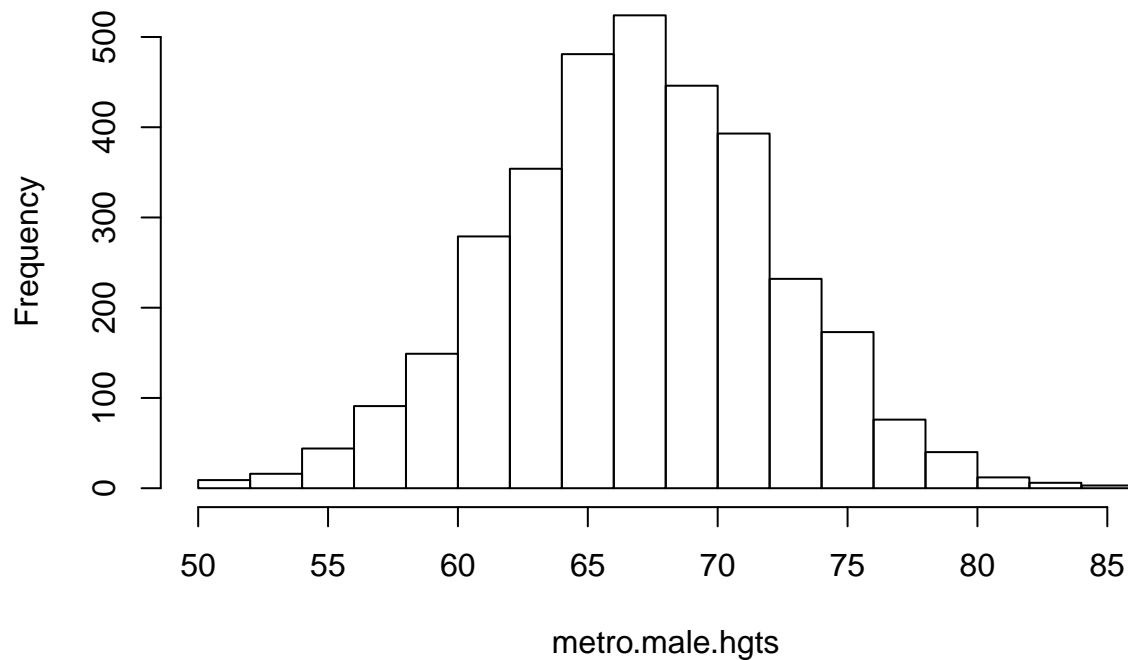
```
metro.male.n <- 3328
metro.male.mu <- 67.5
metro.male.sigma <- 5.25
metro.male.hgts <- round(rnorm(metro.male.n, metro.male.mu, metro.male.sigma))

metro.male.hgts.mean <- mean(metro.male.hgts)
metro.male.hgts.sd <- sd(metro.male.hgts)
metro.male.hgts.mean

## [1] 67.42488
metro.male.hgts.sd

## [1] 5.282839
hist(metro.male.hgts)
```

**Histogram of metro.male.hgts**



```
write.csv(metro.male.hgts, "../metro_hgts_pop.csv", row.names = F)
```

## Samples

```
sample.size <- 30
sample.n <- 100

sample.mat <- matrix(NA, nrow=sample.n, ncol=sample.size)
for (i in 1:sample.n){
  sample.mat[i,] <- sample(metro.male.hgts, sample.size)
}

sample.means <- apply(sample.mat, 1, mean)
sample.sds <- apply(sample.mat, 1, sd)
mean(sample.means)
```

```
## [1] 67.37967
```

```
sd(sample.means)
```

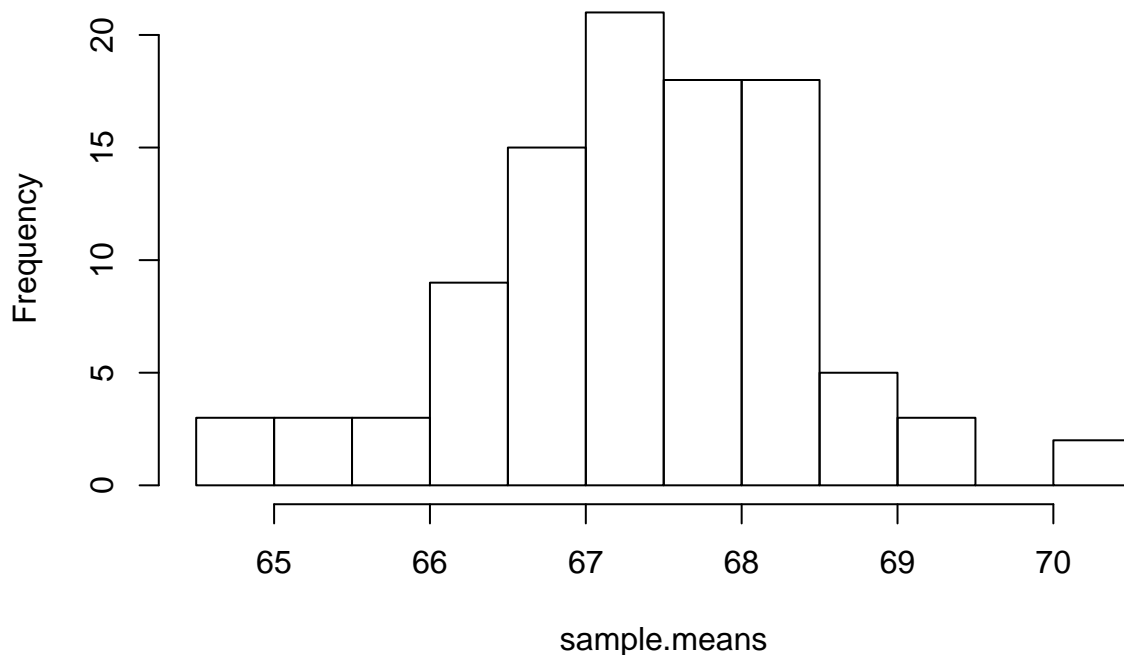
```
## [1] 1.051992
```

```
sd(metro.male.hgts)/sqrt(sample.size)
```

```
## [1] 0.9645101
```

```
hist(sample.means)
```

**Histogram of sample.means**



```
write.csv(sample.mat, "../metro_hgts_samples.csv", row.names = F)
```

```
calc.sample.ci <- function(s){
  s.mean <- mean(s)
  s.sd <- sd(s)
```

```

s.n <- length(s)

s.mean + c(-1,1) * qnorm(0.975) * s.sd/sqrt(s.n)
}

sample.ci <- apply(sample.mat, 1, calc.sample.ci)

sample.ci.df <- data.frame(sample.num=1:sample.n, mean=sample.means, sd=sample.sds,
                           ci.low=sample.ci[1,], ci.up=sample.ci[2,])
sample.ci.df$mean.in.ci <- ifelse(sample.ci.df$ci.low<metro.male.hgts.mean &
                                   sample.ci.df$ci.up>metro.male.hgts.mean,
                                   "yes", "no")

sum(sample.ci.df$mean.in.ci=="yes") / nrow(sample.ci.df)

## [1] 0.9
sample.ci.df$mean[1:5]

## [1] 66.43333 67.16667 68.16667 66.03333 67.23333
sample.ci.df[1,]

##   sample.num    mean      sd  ci.low  ci.up mean.in.ci
## 1           1 66.43333 5.354233 64.51738 68.34929      yes
write.csv(sample.ci.df, "../metro_hgts_sample_stats.csv", row.names = F)

g <- ggplot(sample.ci.df, aes(x=sample.num, y=mean, color=mean.in.ci))
g <- g + geom_hline(yintercept = c(69.2,metro.male.hgts.mean), linetype=2)
g <- g + geom_point( show.legend = F)
g <- g + geom_linerange(aes(ymin=ci.low, ymax=ci.up), size=0.5, show.legend=F)
g <- g + theme_bw()
g <- g + scale_color_manual(values=c(no="red", yes="black"))
g <- g + labs(x="Sample", y="Height (in)")

g

```

