

Assignment 2

Michael Knopf

September 24, 2014

Part A

```
### Read in data, set population size
families = read.csv("/Users/kida/Desktop/Stat 135/Excel CSV/Chapter 7/families.csv")
N = nrow(families)

### Conduct simple random sample
rsamp = families[sample(N, 500), ]

### Create estimated standard error function and confidence interval function
s.xbar = function(x) sqrt(var(x)/length(x) * (1 - length(x)/N))
conf.int = function(x, b)
  {r = qnorm((b+1)/2)*s.xbar(x)
   return (c(mean(x) - r, mean(x) + r))}

### Calculate mean, SE, and confidence for proportion of female-headed families
type = rsamp$TYPE
type[which(type < 3)] = 0
type[which(type == 3)] = 1
femheads.p = mean(type)
femheads.se = s.xbar(type)
femheads.ci = conf.int(type, .95)

### Calculate mean, SE, and confidence for number of children
child.mean = mean(rsamp$CHILDREN)
child.se = s.xbar(rsamp$CHILDREN)
child.ci = conf.int(rsamp$CHILDREN, .95)

### Calculate mean, SE, and confidence for households without diplomas
nodiploma = rsamp$EDUCATION
nodiploma[which(nodiploma <= 38)] = 1
nodiploma[which(nodiploma > 38)] = 0
nodiploma.p = mean(nodiploma)
nodiploma.se = s.xbar(nodiploma)
nodiploma.ci = conf.int(nodiploma, .95)

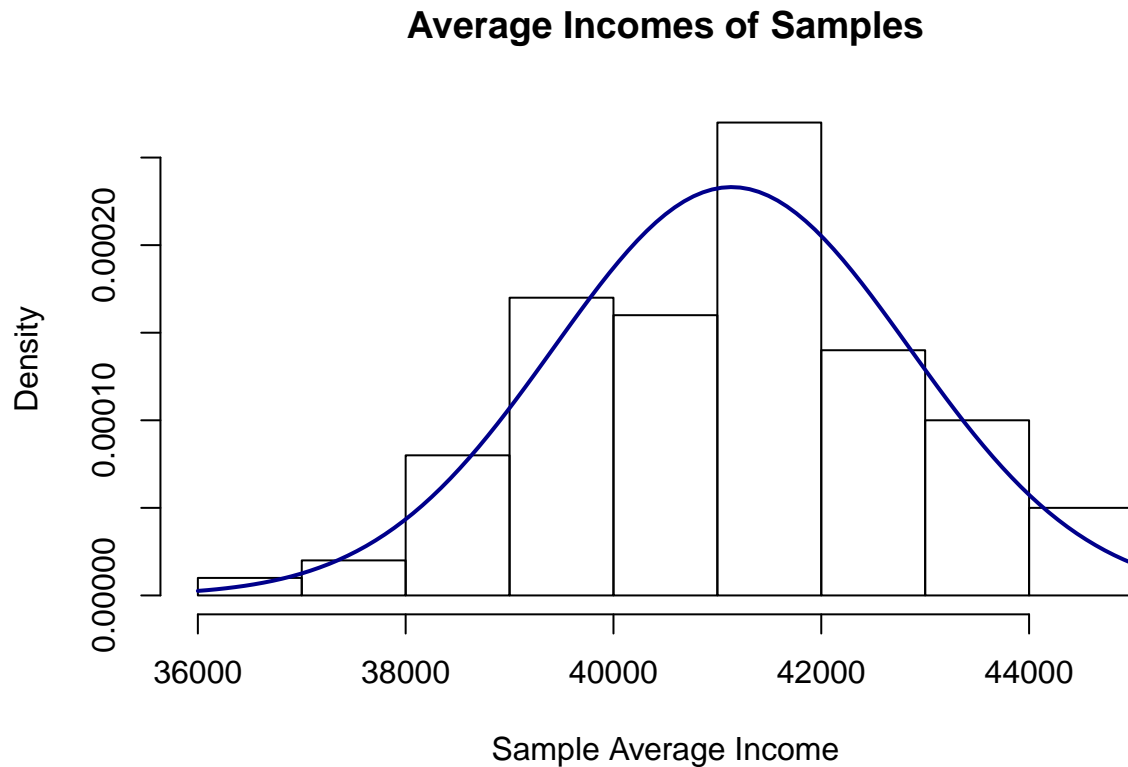
### Calculate mean, SE, and confidence for average family income
income.mean = mean(rsamp$INCOME)
income.se = s.xbar(rsamp$INCOME)
income.ci = conf.int(rsamp$INCOME, .95)

rsamps = list()
for (i in 1:100)
  {rsamps[[i]] = families[sample(N, 400), ]}

avgincomes = sapply(seq(length(rsamps)), function(x) mean(rsamps[[x]]$INCOME))
avgincomes.mean = mean(avgincomes)
```

```
avgincomes.sd = sqrt(99/100)*sd(avgincomes)

hist(avgincomes, freq = FALSE, main = "Average Incomes of Samples",
     xlab = "Sample Average Income")
curve(dnorm(x, mean=avgincomes.mean, sd=avgincomes.sd),
     col="darkblue", lwd=2, add=TRUE, yaxt="n")
```



```
confints = lapply(seq(length(rsamps)), function(x) conf.int(rsamps[[x]]$INCOME, .95))
successes = sapply(seq(length(confints)), function(x)
  (confints[[x]][1] <= mean(families$INCOME) & (mean(families$INCOME) <= confints[[x]][2])))
confidence_proportion = mean(as.numeric(successes))
```