MSDS 6373

UNIT 1

DATA PROFILE – BAR PLOT

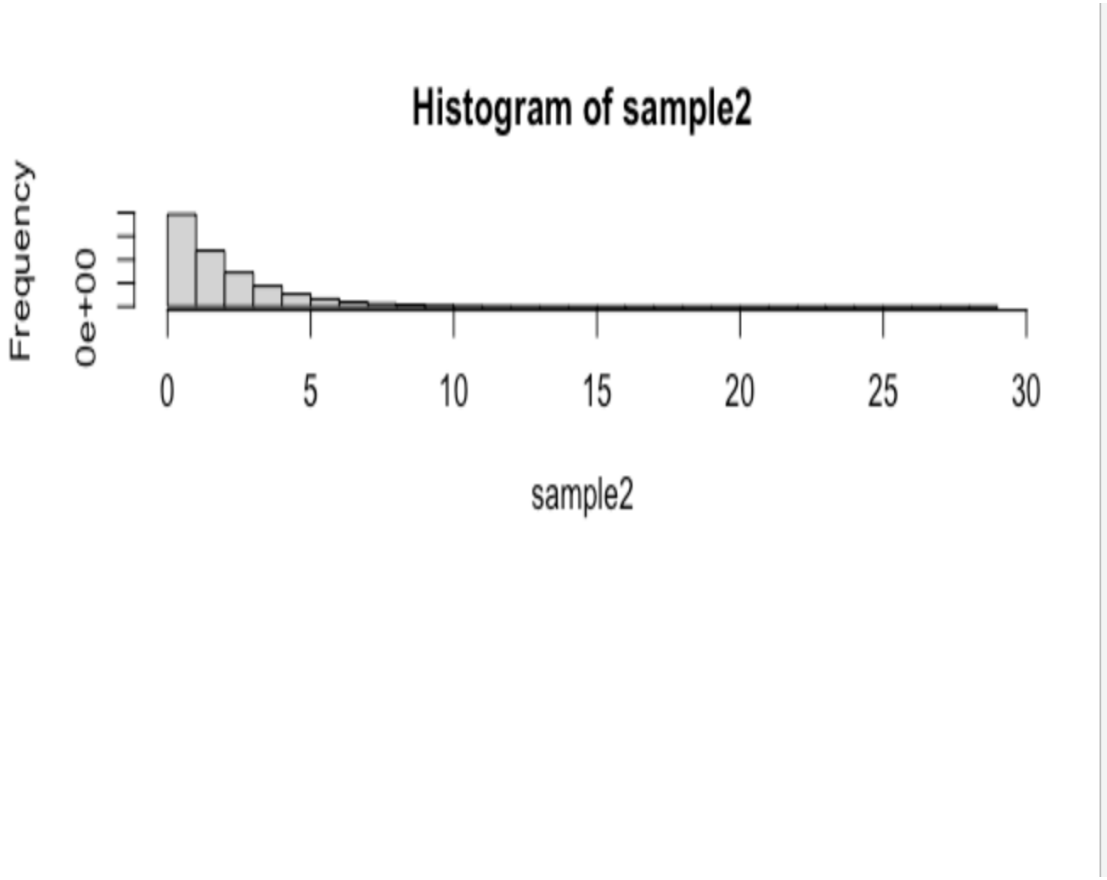
Code

- expdesc = c("Data Vis", "ML", "Stats", "DE", "CS" , "Comm", "Math")
- level = c(5,2,4,3,8,10,7)

- df3 =data.frame(Years = level, Exp = expdesc)
- df3
- barplot(df3$Years, expdesc.arg = df3$Exp, ylab = "Experience",main = "Data Scientist Data Profile")

Unit 1 – Question 2

Histogram



Code, Mean and Standard Deviation of the population

- sample2 = rchisq(10000000,2)
- sample2
- hist(sample2)

- ean(sample2)
- [1] 2.000953
- > sd(sample2)
- [1] 2.001545

Question 4 – Mean/SD according to CLT

- > mean(xbars)
- [1] 1.99587
- > sd(xbars)
- [1] 0.3546334

- mean(xbars)
- [1] 2.001218
- > sd(xbars)
- [1] 0.2823095

Central Limit Theorem Code

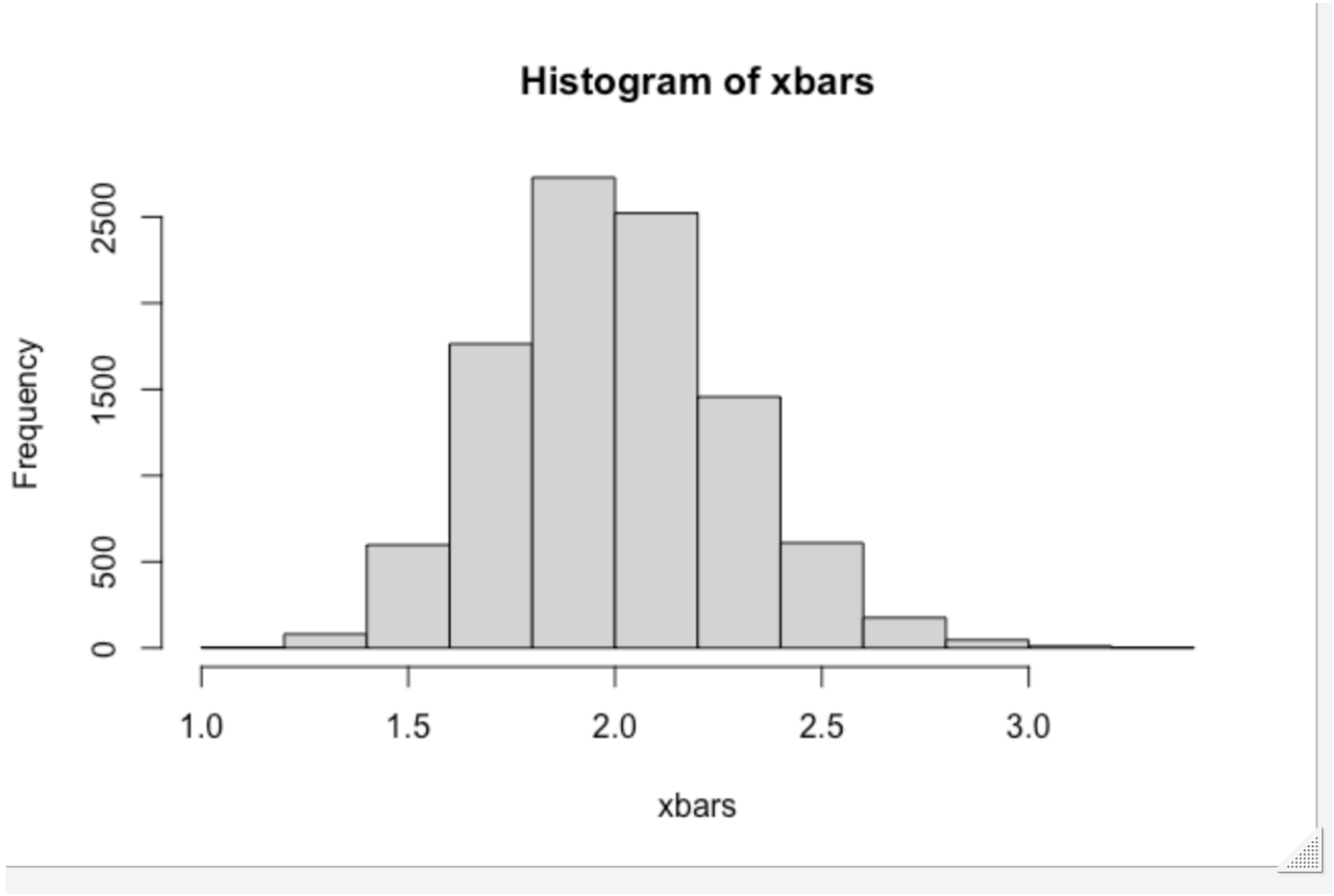
```
xBarVec = c()
population =rchisq(10000000,2)


xbarGenerator = function(sampleSize = 30, number_of_samples = 50)
{
 for(i in 1:number_of_samples)
 {
   theSample =sample(population,sampleSize)
   xbar=mean(theSample)
   xBarVec = c(xBarVec, xbar)
 }
 return(xBarVec)
}

 xbars = xbarGenerator(30,1000)
 hist(xbars)
mean(xbars)
sd(xbars)
```

T-Test

Question 5/6



- mean(xbars)
- [1] 2.001218
- > sd(xbars)
- [1] 0.2823095

Takeaways and Questions

- 1)The screens are extremely blurry so had to pause several times to get the details for the CLT code. I tried it on several computers.
- 2) Bar Plot - I used names.arg but the names didn’t print on the bar plot.
- 3)I couldn’t find the details for the T-Test
- 4) I found it extremely helpful to work through the exercise and write the code. It allowed me to understand better. thanks