HW 1

Ishita Dutta

4/14/2022

Part 2)

a)

```
pop <- c(15, 34, 35, 36, 11, 17, 36, 15)
samples <- c()
sums <- NULL
sum_count <- NULL
sumval <- NULL
n <- 3
N <- 8

for (i in 1:8){
  for (j in 1:8){
    for (k in 1:8){
      sample <- c( pop[i], pop[j],pop[k] )
      samples <- rbind(samples, sample)
      sval = pop[i] + pop[j] + pop[k]
      if (sval %in% sumval){
        index = match(sval, sumval)
        sum_count[index] = sum_count[index] + 1
      }
      else {
        sumval = c(sumval, sval)
        sum_count = c(sum_count, 1)
      }
    }
  }
}
samples
```

```
##        [,1] [,2] [,3]
## sample   15   15   15
## sample   15   15   34
## sample   15   15   35
## sample   15   15   36
## sample   15   15   11
## sample   15   15   17
## sample   15   15   36
```

```
## sample   15   15   15
## sample   15   34   15
## sample   15   34   34
## sample   15   34   35
## sample   15   34   36
## sample   15   34   11
## sample   15   34   17
## sample   15   34   36
## sample   15   34   15
## sample   15   35   15
## sample   15   35   34
## sample   15   35   35
## sample   15   35   36
## sample   15   35   11
## sample   15   35   17
## sample   15   35   36
## sample   15   35   15
## sample   15   36   15
## sample   15   36   34
## sample   15   36   35
## sample   15   36   36
## sample   15   36   11
## sample   15   36   17
## sample   15   36   36
## sample   15   36   15
## sample   15   11   15
## sample   15   11   34
## sample   15   11   35
## sample   15   11   36
## sample   15   11   11
## sample   15   11   17
## sample   15   11   36
## sample   15   11   15
## sample   15   17   15
## sample   15   17   34
## sample   15   17   35
## sample   15   17   36
## sample   15   17   11
## sample   15   17   17
## sample   15   17   36
## sample   15   17   15
## sample   15   36   15
## sample   15   36   34
## sample   15   36   35
## sample   15   36   36
## sample   15   36   11
## sample   15   36   17
## sample   15   36   36
## sample   15   36   15
## sample   15   15   15
## sample   15   15   34
## sample   15   15   35
## sample   15   15   36
## sample   15   15   11
```

```
## sample   15   15   17
## sample   15   15   36
## sample   15   15   15
## sample   34   15   15
## sample   34   15   34
## sample   34   15   35
## sample   34   15   36
## sample   34   15   11
## sample   34   15   17
## sample   34   15   36
## sample   34   15   15
## sample   34   34   15
## sample   34   34   34
## sample   34   34   35
## sample   34   34   36
## sample   34   34   11
## sample   34   34   17
## sample   34   34   36
## sample   34   34   15
## sample   34   35   15
## sample   34   35   34
## sample   34   35   35
## sample   34   35   36
## sample   34   35   11
## sample   34   35   17
## sample   34   35   36
## sample   34   35   15
## sample   34   36   15
## sample   34   36   34
## sample   34   36   35
## sample   34   36   36
## sample   34   36   11
## sample   34   36   17
## sample   34   36   36
## sample   34   36   15
## sample   34   11   15
## sample   34   11   34
## sample   34   11   35
## sample   34   11   36
## sample   34   11   11
## sample   34   11   17
## sample   34   11   36
## sample   34   11   15
## sample   34   17   15
## sample   34   17   34
## sample   34   17   35
## sample   34   17   36
## sample   34   17   11
## sample   34   17   17
## sample   34   17   36
## sample   34   17   15
## sample   34   36   15
## sample   34   36   34
## sample   34   36   35
```

```
## sample   34   36   36
## sample   34   36   11
## sample   34   36   17
## sample   34   36   36
## sample   34   36   15
## sample   34   15   15
## sample   34   15   34
## sample   34   15   35
## sample   34   15   36
## sample   34   15   11
## sample   34   15   17
## sample   34   15   36
## sample   34   15   15
## sample   35   15   15
## sample   35   15   34
## sample   35   15   35
## sample   35   15   36
## sample   35   15   11
## sample   35   15   17
## sample   35   15   36
## sample   35   15   15
## sample   35   34   15
## sample   35   34   34
## sample   35   34   35
## sample   35   34   36
## sample   35   34   11
## sample   35   34   17
## sample   35   34   36
## sample   35   34   15
## sample   35   35   15
## sample   35   35   34
## sample   35   35   35
## sample   35   35   36
## sample   35   35   11
## sample   35   35   17
## sample   35   35   36
## sample   35   35   15
## sample   35   36   15
## sample   35   36   34
## sample   35   36   35
## sample   35   36   36
## sample   35   36   11
## sample   35   36   17
## sample   35   36   36
## sample   35   36   15
## sample   35   11   15
## sample   35   11   34
## sample   35   11   35
## sample   35   11   36
## sample   35   11   11
## sample   35   11   17
## sample   35   11   36
## sample   35   11   15
## sample   35   17   15
```

```
## sample   35   17   34
## sample   35   17   35
## sample   35   17   36
## sample   35   17   11
## sample   35   17   17
## sample   35   17   36
## sample   35   17   15
## sample   35   36   15
## sample   35   36   34
## sample   35   36   35
## sample   35   36   36
## sample   35   36   11
## sample   35   36   17
## sample   35   36   36
## sample   35   36   15
## sample   35   15   15
## sample   35   15   34
## sample   35   15   35
## sample   35   15   36
## sample   35   15   11
## sample   35   15   17
## sample   35   15   36
## sample   35   15   15
## sample   36   15   15
## sample   36   15   34
## sample   36   15   35
## sample   36   15   36
## sample   36   15   11
## sample   36   15   17
## sample   36   15   36
## sample   36   15   15
## sample   36   34   15
## sample   36   34   34
## sample   36   34   35
## sample   36   34   36
## sample   36   34   11
## sample   36   34   17
## sample   36   34   36
## sample   36   34   15
## sample   36   35   15
## sample   36   35   34
## sample   36   35   35
## sample   36   35   36
## sample   36   35   11
## sample   36   35   17
## sample   36   35   36
## sample   36   35   15
## sample   36   36   15
## sample   36   36   34
## sample   36   36   35
## sample   36   36   36
## sample   36   36   11
## sample   36   36   17
## sample   36   36   36
```

```
## sample   36   36   15
## sample   36   11   15
## sample   36   11   34
## sample   36   11   35
## sample   36   11   36
## sample   36   11   11
## sample   36   11   17
## sample   36   11   36
## sample   36   11   15
## sample   36   17   15
## sample   36   17   34
## sample   36   17   35
## sample   36   17   36
## sample   36   17   11
## sample   36   17   17
## sample   36   17   36
## sample   36   17   15
## sample   36   36   15
## sample   36   36   34
## sample   36   36   35
## sample   36   36   36
## sample   36   36   11
## sample   36   36   17
## sample   36   36   36
## sample   36   36   15
## sample   36   15   15
## sample   36   15   34
## sample   36   15   35
## sample   36   15   36
## sample   36   15   11
## sample   36   15   17
## sample   36   15   36
## sample   36   15   15
## sample   11   15   15
## sample   11   15   34
## sample   11   15   35
## sample   11   15   36
## sample   11   15   11
## sample   11   15   17
## sample   11   15   36
## sample   11   15   15
## sample   11   34   15
## sample   11   34   34
## sample   11   34   35
## sample   11   34   36
## sample   11   34   11
## sample   11   34   17
## sample   11   34   36
## sample   11   34   15
## sample   11   35   15
## sample   11   35   34
## sample   11   35   35
## sample   11   35   36
## sample   11   35   11
```

```
## sample   11   35   17
## sample   11   35   36
## sample   11   35   15
## sample   11   36   15
## sample   11   36   34
## sample   11   36   35
## sample   11   36   36
## sample   11   36   11
## sample   11   36   17
## sample   11   36   36
## sample   11   36   15
## sample   11   11   15
## sample   11   11   34
## sample   11   11   35
## sample   11   11   36
## sample   11   11   11
## sample   11   11   17
## sample   11   11   36
## sample   11   11   15
## sample   11   17   15
## sample   11   17   34
## sample   11   17   35
## sample   11   17   36
## sample   11   17   11
## sample   11   17   17
## sample   11   17   36
## sample   11   17   15
## sample   11   36   15
## sample   11   36   34
## sample   11   36   35
## sample   11   36   36
## sample   11   36   11
## sample   11   36   17
## sample   11   36   36
## sample   11   36   15
## sample   11   15   15
## sample   11   15   34
## sample   11   15   35
## sample   11   15   36
## sample   11   15   11
## sample   11   15   17
## sample   11   15   36
## sample   11   15   15
## sample   17   15   15
## sample   17   15   34
## sample   17   15   35
## sample   17   15   36
## sample   17   15   11
## sample   17   15   17
## sample   17   15   36
## sample   17   15   15
## sample   17   34   15
## sample   17   34   34
## sample   17   34   35
```

```
## sample   17   34   36
## sample   17   34   11
## sample   17   34   17
## sample   17   34   36
## sample   17   34   15
## sample   17   35   15
## sample   17   35   34
## sample   17   35   35
## sample   17   35   36
## sample   17   35   11
## sample   17   35   17
## sample   17   35   36
## sample   17   35   15
## sample   17   36   15
## sample   17   36   34
## sample   17   36   35
## sample   17   36   36
## sample   17   36   11
## sample   17   36   17
## sample   17   36   36
## sample   17   36   15
## sample   17   11   15
## sample   17   11   34
## sample   17   11   35
## sample   17   11   36
## sample   17   11   11
## sample   17   11   17
## sample   17   11   36
## sample   17   11   15
## sample   17   17   15
## sample   17   17   34
## sample   17   17   35
## sample   17   17   36
## sample   17   17   11
## sample   17   17   17
## sample   17   17   36
## sample   17   17   15
## sample   17   36   15
## sample   17   36   34
## sample   17   36   35
## sample   17   36   36
## sample   17   36   11
## sample   17   36   17
## sample   17   36   36
## sample   17   36   15
## sample   17   15   15
## sample   17   15   34
## sample   17   15   35
## sample   17   15   36
## sample   17   15   11
## sample   17   15   17
## sample   17   15   36
## sample   17   15   15
## sample   36   15   15
```

```
## sample   36   15   34
## sample   36   15   35
## sample   36   15   36
## sample   36   15   11
## sample   36   15   17
## sample   36   15   36
## sample   36   15   15
## sample   36   34   15
## sample   36   34   34
## sample   36   34   35
## sample   36   34   36
## sample   36   34   11
## sample   36   34   17
## sample   36   34   36
## sample   36   34   15
## sample   36   35   15
## sample   36   35   34
## sample   36   35   35
## sample   36   35   36
## sample   36   35   11
## sample   36   35   17
## sample   36   35   36
## sample   36   35   15
## sample   36   36   15
## sample   36   36   34
## sample   36   36   35
## sample   36   36   36
## sample   36   36   11
## sample   36   36   17
## sample   36   36   36
## sample   36   36   15
## sample   36   11   15
## sample   36   11   34
## sample   36   11   35
## sample   36   11   36
## sample   36   11   11
## sample   36   11   17
## sample   36   11   36
## sample   36   11   15
## sample   36   17   15
## sample   36   17   34
## sample   36   17   35
## sample   36   17   36
## sample   36   17   11
## sample   36   17   17
## sample   36   17   36
## sample   36   17   15
## sample   36   36   15
## sample   36   36   34
## sample   36   36   35
## sample   36   36   36
## sample   36   36   11
## sample   36   36   17
## sample   36   36   36
```

```
## sample   36   36   15
## sample   36   15   15
## sample   36   15   34
## sample   36   15   35
## sample   36   15   36
## sample   36   15   11
## sample   36   15   17
## sample   36   15   36
## sample   36   15   15
## sample   15   15   15
## sample   15   15   34
## sample   15   15   35
## sample   15   15   36
## sample   15   15   11
## sample   15   15   17
## sample   15   15   36
## sample   15   15   15
## sample   15   34   15
## sample   15   34   34
## sample   15   34   35
## sample   15   34   36
## sample   15   34   11
## sample   15   34   17
## sample   15   34   36
## sample   15   34   15
## sample   15   35   15
## sample   15   35   34
## sample   15   35   35
## sample   15   35   36
## sample   15   35   11
## sample   15   35   17
## sample   15   35   36
## sample   15   35   15
## sample   15   36   15
## sample   15   36   34
## sample   15   36   35
## sample   15   36   36
## sample   15   36   11
## sample   15   36   17
## sample   15   36   36
## sample   15   36   15
## sample   15   11   15
## sample   15   11   34
## sample   15   11   35
## sample   15   11   36
## sample   15   11   11
## sample   15   11   17
## sample   15   11   36
## sample   15   11   15
## sample   15   17   15
## sample   15   17   34
## sample   15   17   35
## sample   15   17   36
## sample   15   17   11
```

```
## sample   15   17   17
## sample   15   17   36
## sample   15   17   15
## sample   15   36   15
## sample   15   36   34
## sample   15   36   35
## sample   15   36   36
## sample   15   36   11
## sample   15   36   17
## sample   15   36   36
## sample   15   36   15
## sample   15   15   15
## sample   15   15   34
## sample   15   15   35
## sample   15   15   36
## sample   15   15   11
## sample   15   15   17
## sample   15   15   36
## sample   15   15   15
```

b)

```r
n = sum(sum_count)
prob_vals = sum_count / n
sum_probs = data.frame(sumval, sum_count, prob_vals)
sum_probs
```

```
##    sumval sum_count   prob_vals
## 1      45        11 0.021484375
## 2      64        24 0.046875000
## 3      65        12 0.023437500
## 4      66        36 0.070312500
## 5      41        12 0.023437500
## 6      47        12 0.023437500
## 7      83        18 0.035156250
## 8      84        12 0.023437500
## 9      85        33 0.064453125
## 10     60        12 0.023437500
## 11     86        30 0.058593750
## 12     61        12 0.023437500
## 13     67        12 0.023437500
## 14     87        39 0.076171875
## 15     62        30 0.058593750
## 16     68        27 0.052734375
## 17     37         6 0.011718750
## 18     43        12 0.023437500
## 19     49         6 0.011718750
## 20    102         1 0.001953125
## 21    103         3 0.005859375
## 22    104         9 0.017578125
## 23     79         3 0.005859375
## 24    105        13 0.025390625
```

```
## 25      80          6 0.011718750
## 26     106         18 0.035156250
## 27      81         15 0.029296875
## 28      56          3 0.005859375
## 29     107         12 0.023437500
## 30      82         12 0.023437500
## 31      88         12 0.023437500
## 32      57          3 0.005859375
## 33      63          6 0.011718750
## 34      69          3 0.005859375
## 35     108          8 0.015625000
## 36      89         12 0.023437500
## 37      58          6 0.011718750
## 38      70          6 0.011718750
## 39      33          1 0.001953125
## 40      39          3 0.005859375
## 41      51          1 0.001953125
```

c)

```
n = sum(sum_probs$sum_count)
z1 = 1/n
z0 = (n-1)/n
Z_vals = c(0,1)
Z_probs = c(z0, z1)
eZ = sum(Z_vals * Z_probs)
varZ = sum(((Z_vals - eZ)^2) * Z_probs )

z00 =z0^2
z10 = z1*z0
z01 = z0*z1
z11 = z1^2
probszizj = c(z00, z01, z10, z11)
covZ = 0
for(i in c(1:2)) {
  for (j in  c(1:2)) {
    X = Z_vals[i]
    Y = Z_vals[j]
    covZ = covZ + ((Y - eZ)*(X - eZ)*(probszizj[i + j]))
  }
}
corrZ = covZ/varZ
corrZ
```

```
## [1] -0.001945496
```

d)

```
nsums <- NULL
nsum_count <- NULL
```

```r
nsumval <- NULL
for(i in c(1:8)) {
  for(j in c(1:7)) {
    for(k in c(1:6)) {
      nsval = pop[i] + pop[j] + pop[k]
      #print(c(i,j, k))
      if (nsval %in% nsumval){
        index = match(nsval, nsumval)
        nsum_count[index] = nsum_count[index] + 1
      }
      else {
        nsumval = c(nsumval, nsval)
        nsum_count = c(nsum_count, 1)
      }
    }
  }
}
m = sum(nsum_count)
nprob_vals = nsum_count/m
nsum_probs = data.frame(nsumval, nsum_count, nprob_vals)
nsum_probs
```

```
##    nsumval nsum_count  nprob_vals
## 1       45          5 0.014880952
## 2       64         15 0.044642857
## 3       65          5 0.014880952
## 4       66         16 0.047619048
## 5       41          5 0.014880952
## 6       47          5 0.014880952
## 7       83         12 0.035714286
## 8       84          8 0.023809524
## 9       85         20 0.059523810
## 10      60          8 0.023809524
## 11      86         19 0.056547619
## 12      61          8 0.023809524
## 13      67          8 0.023809524
## 14      87         23 0.068452381
## 15      62         19 0.056547619
## 16      68         16 0.047619048
## 17      37          4 0.011904762
## 18      43          8 0.023809524
## 19      49          4 0.011904762
## 20     102          1 0.002976190
## 21     103          3 0.008928571
## 22     104          8 0.023809524
## 23      79          3 0.008928571
## 24     105         11 0.032738095
## 25      80          6 0.017857143
## 26     106         13 0.038690476
## 27      81         13 0.038690476
## 28      56          3 0.008928571
## 29     107          8 0.023809524
## 30      82         10 0.029761905
```

```
## 31      88         10 0.029761905
## 32      57          3 0.008928571
## 33      63          6 0.017857143
## 34      69          3 0.008928571
## 35     108          4 0.011904762
## 36      89          8 0.023809524
## 37      58          5 0.014880952
## 38      70          5 0.014880952
## 39      33          1 0.002976190
## 40      39          3 0.008928571
## 41      51          1 0.002976190
```

e)

```
nz1 = 1/m
nz0 = (m-1)/m
nZ_vals = c(0,1)
nZ_probs = c(nz0, nz1)
neZ = sum(nZ_vals * nZ_probs)
nvarZ = sum(((nZ_vals - neZ)^2) * nZ_probs )

nz00 =nz0^2
nz10 = nz1*z0
nz01 = nz0*nz1
nz11 = nz1^2
nprobszizj = c(nz00, nz01, nz10, nz11)
ncovZ = 0
for(i in c(1:2)) {
  for (j in  c(1:2)) {
    X = nZ_vals[i]
    Y = nZ_vals[j]
    ncovZ = ncovZ + ((Y - eZ)*(X - eZ)*(probszizj[i + j]))
  }
}
ncorrZ = ncovZ/nvarZ
ncorrZ
```

```
## [1] -0.001278042
```

f)

```
# with replacement
e = sum(sum_probs$sumval * sum_probs$prob_vals)
vari = sum(((sum_probs$sumval - e)^2) * sum_probs$prob_vals)
cat("With Replacement:\n Mean:", e, "\nSE:", vari, "\n\n")
```

```
## With Replacement:
##  Mean: 74.625
## SE: 331.0781
```

```r
# without replacement
ne = sum(nsum_probs$nsumval * nsum_probs$nprob_vals)
nvari = sum(((nsum_probs$nsumval - ne)^2) * nsum_probs$nprob_vals)
cat("Without Replacement:\n Mean:", ne, "\nSE:", nvari)
```

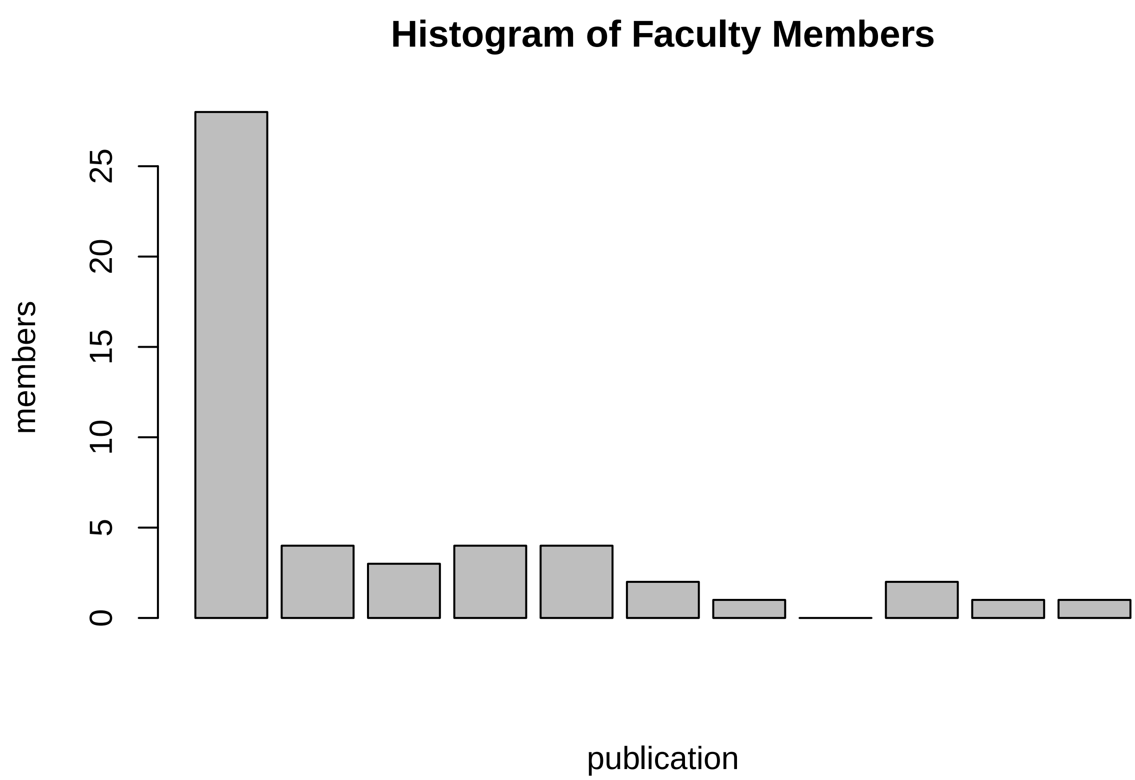
```
## Without Replacement:
##  Mean: 75.82738
## SE: 330.7857
```

Part 3

a)

```r
members=c(28,4,3,4,4,2,1,0,2,1,1)
publication = c(0,1,2,3,4,5,6,7,8,9,10)

barplot(members, main = "Histogram of Faculty Members",ylab = "members",xlab = "publication")
```



Solution: Shape of the data is very right skewed.

b)

```r
#member_mean = ((28*0)+(1*4)+(2*3)+(3*4)+(4*4)+(5*2)+(6*1)+(7*0)+(8*2)+(9*1)+(10*1))/50
#member_mean
n = 50
N = 807
weight <- rep(N/n, n)
ans <- rep(publication, members)
index <- c(1:50)
published_faculty <- data.frame(index,ans)


published_design <- svydesign(id = ~1, weights = weight, fpc = rep(N,n), data = published_faculty)
ans <- svymean(~ans, published_design)
ans
```

```
##     mean     SE
## ans 1.78 0.3674
```

```r
###
```

c)

Solution: No, since the distribution is right-skewed and SE is 0.3674, it is not right to define that the distribution is normally distributed.

d)

```r
# proportion of faculty members with no publications and give a 95% CI.
p <- 28/n


se <- sqrt(((p*(1-p))/n)*(1-n/N))
se
```

```
## [1] 0.06799023
```

```r
conf_int_pos <- p + (1.96*se)
conf_int_neg <- p - (1.96*se)

print(c(conf_int_neg, conf_int_pos))
```

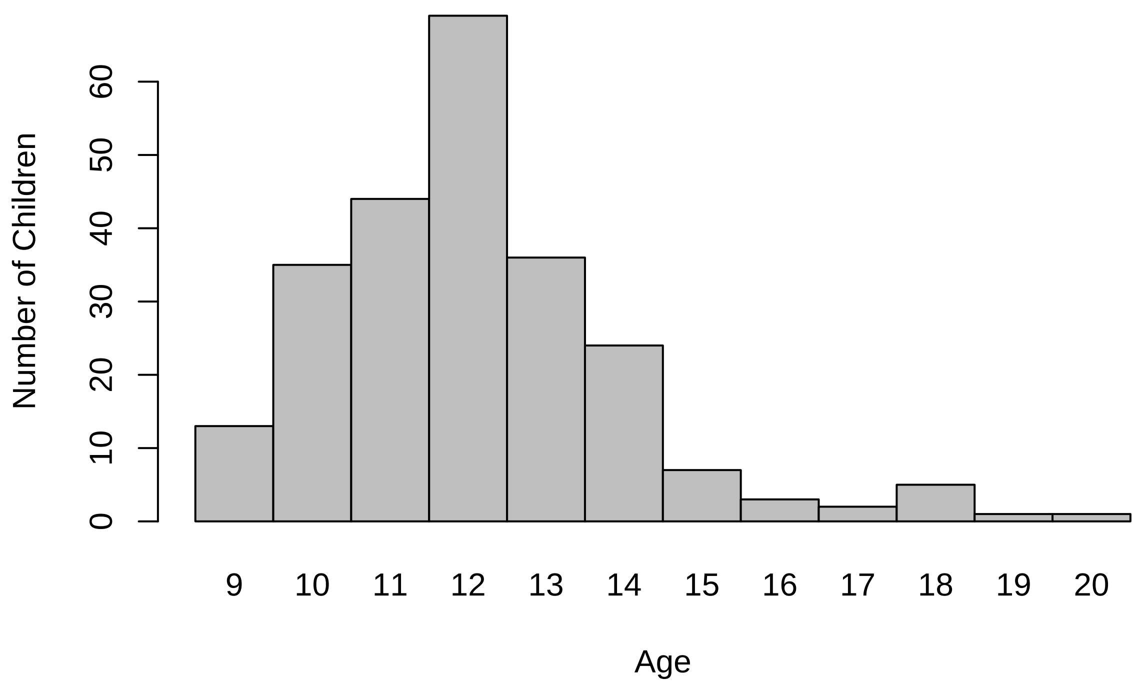
```
## [1] 0.4267391 0.6932609
```

Solution: 95% C.I is [0.4267391 0.6932609]

Part 4)

a)

```
n = 240
age = c(9:20)
num_children = c(13, 35, 44, 69, 36, 24, 7, 3, 2, 5, 1, 1)
tab = data.frame(age, num_children)
barplot(tab$num_children, names.arg=tab$age, ylab="Number of Children", xlab="Age", space = 0)
```



Solution: The shape is not exactly normally distributed. It has a slight right skew in the histogram above. In this case, we would still assume normality by applying law of large numbers.

b)

```
# Getting probabilities of each value
tab$prob = round(tab$num_children/n, 4)

# mean
tab.mean = sum(tab$age * tab$prob)
cat("Mean:" , tab.mean, "\n")
```

```
## Mean: 12.0794
```

```r
# SE
vari = sum((tab$age^2) * tab$prob) - (tab.mean ^ 2)
tab.se = sqrt(vari)/sqrt(240)
cat("SE:" , tab.se, "\n")
```

```
## SE: 0.1240245
```

```r
# CI
lower = tab.mean - (tab.se * 1.96)
upper = tab.mean + (tab.se * 1.96)
qnorm(0.975)
```

```
## [1] 1.959964
```

```r
c(lower, upper)
```

```
## [1] 11.83631 12.32249
```

c)

```r
n0 = ((1.96 * sqrt(vari))^2)/(0.5^2)
valn = (n0)/(1 + (n0/240))
round(valn, 0)
```

```
## [1] 46
```

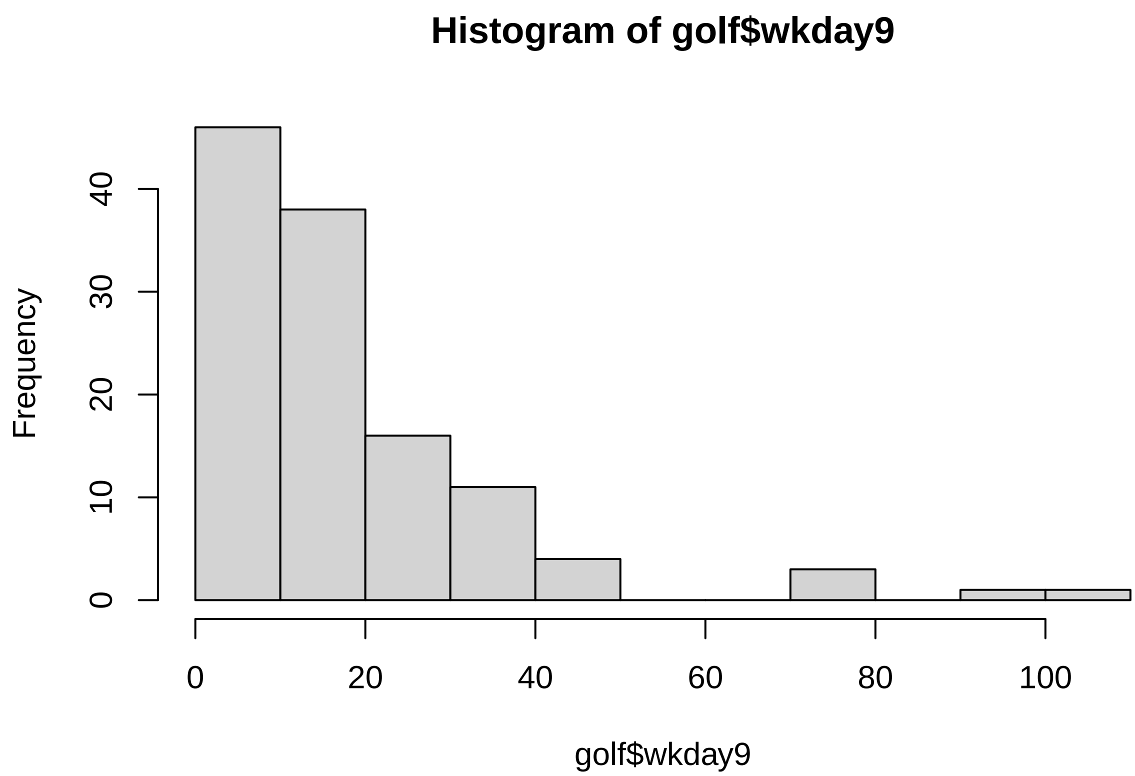
Part 5)

a)

```r
golf = read.csv("golfsrs.csv", header = TRUE)
# print first 5 vals to show we have the table
head(golf)
```

```
##      RN state holes type yearblt wkday18 wkday9 wkend18 wkend9 backtee rating
## 1  5491    RI    18 priv    1923      25     25      35     25    6453   71.8
## 2 10276    VT    18 semi    1972      40     24      45     24    6549   71.1
## 3  6025    MN     9  pub    1939      NA     10      NA     10    3058   69.2
## 4  9739    GA    18 semi    1991      37     37      45     37    6766   72.2
## 5  3463    CA    18  pub    1970      17     10      20     10    6706   71.4
## 6  5883    MN    18  pub    1996      16     12      18     12    7002   73.5
##   par cart18 cart9 caddy pro
## 1  69   15.0   7.5     y   y
## 2  72   30.0  18.0     n   y
## 3  35   16.5  11.0     n   n
## 4  72    0.0   0.0     n   y
## 5  72   22.0  15.0     n   y
## 6  72   10.0   7.0     n   y
```

```r
# getting the histogram of wkday9
hist(golf$wkday9)
```



Solution: I would describe this as the graph for the exponential distribution.

b)

```r
xbar = mean(golf$wkday9)
cat("Mean:", round(xbar, 4), "\n")
```

```
## Mean: 20.1533
```

```r
se = sd(golf$wkday9) * (length(golf$wkday9)/(length(golf$wkday9) - 1))
cat("SE:", round(se, 4), "\n")
```

```
## SE: 18.0771
```

Part 6)

b)

Well, I don't have time to calculate this but smaller values of p generally need higher sample sizes.