Lab3

Omkar Bhutra

2 February 2019

Question 1:

Cluster Sampling

An opinion pool is assumed to be performed in several locations of Sweden by sending interviewers to this location. Of course, it is unreasonable from the fnancial point of view to visit each city. Instead, a decision was done to use random sampling without replacement with the probabilities proportional to the number of inhabitants of the city to select 20 cities. Explore the file population.xls. Note that names in bold are counties, not cities.

2. Use a uniform random number generator to create a function that selects 1 city from the whole list by the probability scheme ordered above (do not use standard sampling functions present in R).

```
## [1] "Ronneby"
```

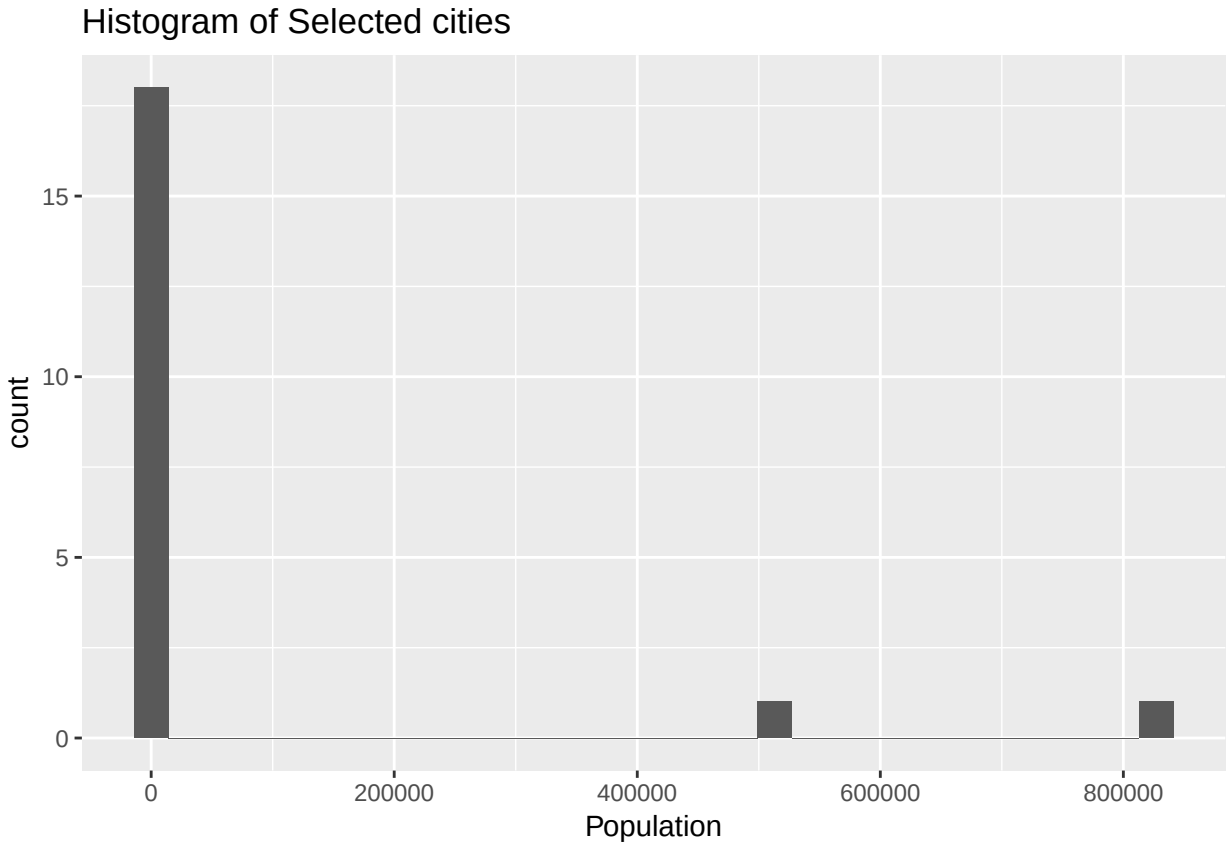
3. . Use the function you have created in step 2 as follows:

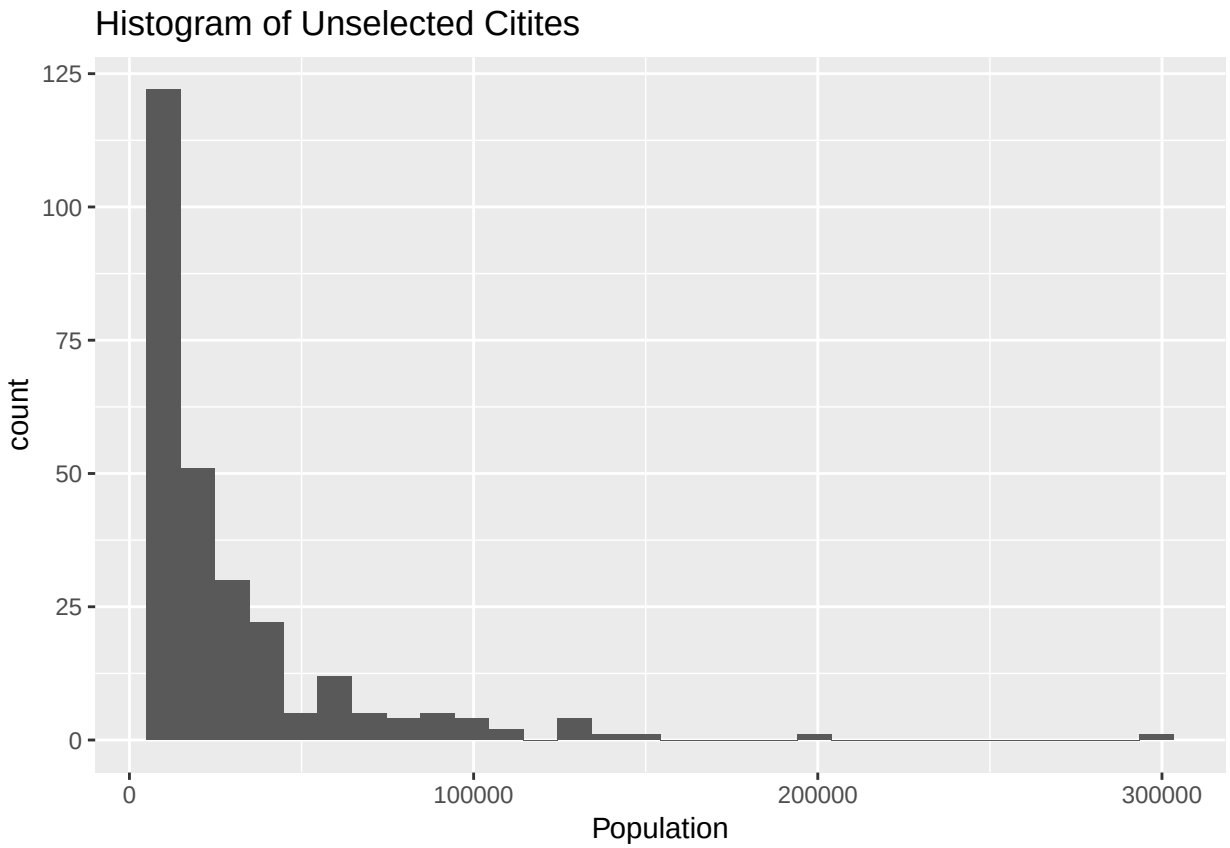
(a) Apply it to the list of all cities and select one city
(b) Remove this city from the list
(c) Apply this function again to the updated list of the cities
(d) Remove this city from the list
(e) ... and so on until you get exactly 20 cities.

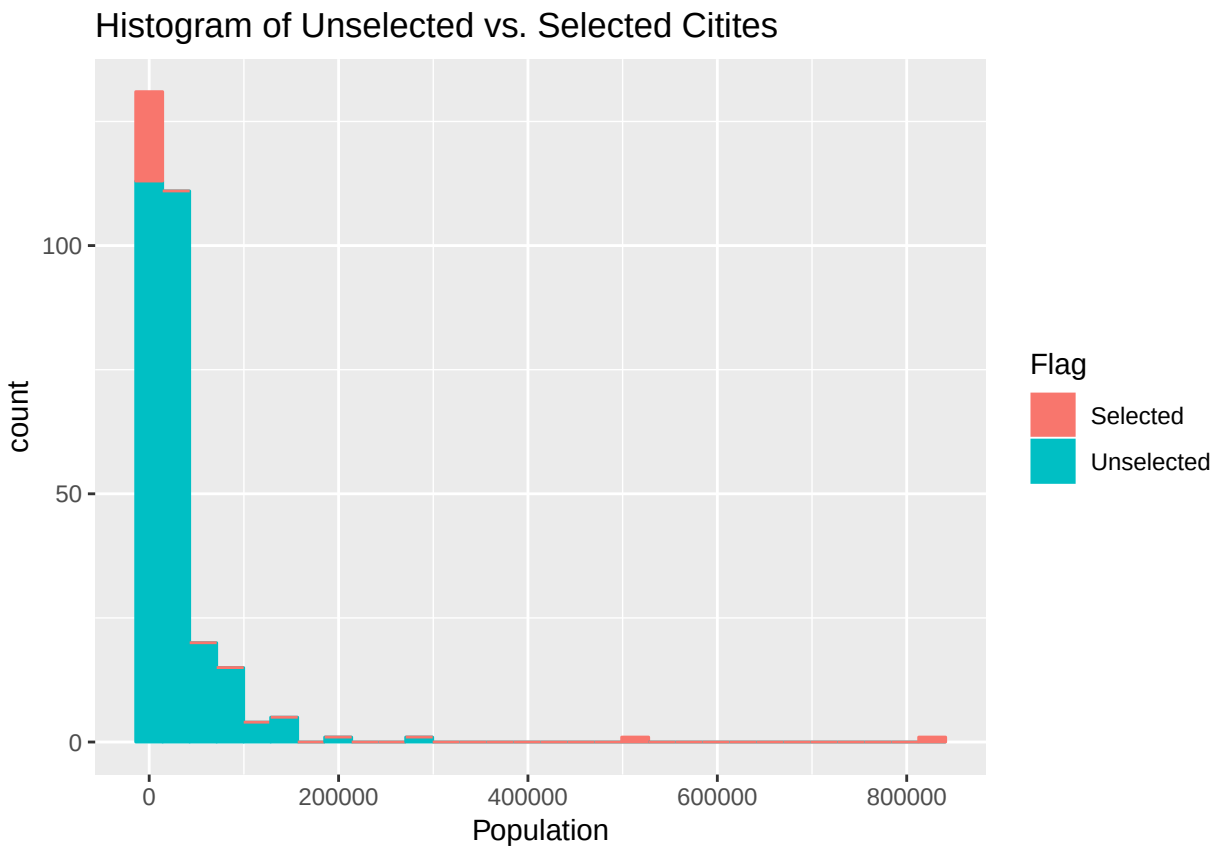
```
## [1] "Kävlinge"
```

4. Run the program. Which cities were selected? What can you say about the size of the selected cities?
5. Plot one histogram showing the size of all cities of the country. Plot another histogram showing the size of the 20 selected cities. Conclusions?

```
##  [1] "Stockholm"       "Boxholm"         "Ydre"
##  [4] "Ödeshög"         "Dals-Ed"         "Göteborg"
##  [7] "Munkfors"        "Storfors"        "Ljusnarsberg"
## [10] "Skinnskatteberg" "Bjurholm"        "Dorotea"
## [13] "Malå"            "Norsjö"          "Sorsele"
## [16] "Åsele"           "Arjeplog"        "Jokkmokk"
## [19] "Överkalix"       "Övertorneå"
```


Histogram of Selected cities
count
15
10
5
0
0
200000
400000
600000
800000
Population


Histogram of Unselected Citites
125
100
75
50
25
0
count
0
100000
200000
300000
Population



Question 2: Different distributions

The double exponential (Laplace) distribution is given by formula:

$$DE(\mu, \alpha) = \frac{\alpha}{2} e^{(-\alpha|x-\mu|)}$$

The CDF is given by:

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$F(x) = \int_{-\infty}^{x} \frac{\alpha}{2} e^{-\alpha(x-\mu)} dx, \qquad (if \quad x > \mu)$$

$$= 1 - \int_{x}^{\infty} \frac{\alpha}{2} e^{-\alpha(x-\mu)} dx$$

$$= 1 - \frac{1}{2} e^{-\alpha(x-\mu)}$$

$$F(x) = \int_{-\infty}^{x} \frac{\alpha}{2} e^{\alpha(x-\mu)} dx, \qquad (if \quad x \leq \mu)$$

$$= \frac{1}{2} e^{\alpha(x-\mu)}$$

Inverse of CDF

$$For \quad x > \mu, \ we \ got \ F(x) = 1 - \frac{1}{2}e^{-\alpha(x-\mu)}$$

$$y = 1 - \frac{1}{2}e^{-\alpha(x-\mu)}$$

$$\frac{ln(2-2y) - \alpha\mu}{-\alpha} = x$$

$$For \ U \sim U(0,1), \qquad \frac{ln(2-2U) - \alpha\mu}{-\alpha} = X$$

$$For \quad x \leq \mu, \ we \ got \ F(x) = \frac{1}{2}e^{\alpha(x-\mu)}$$

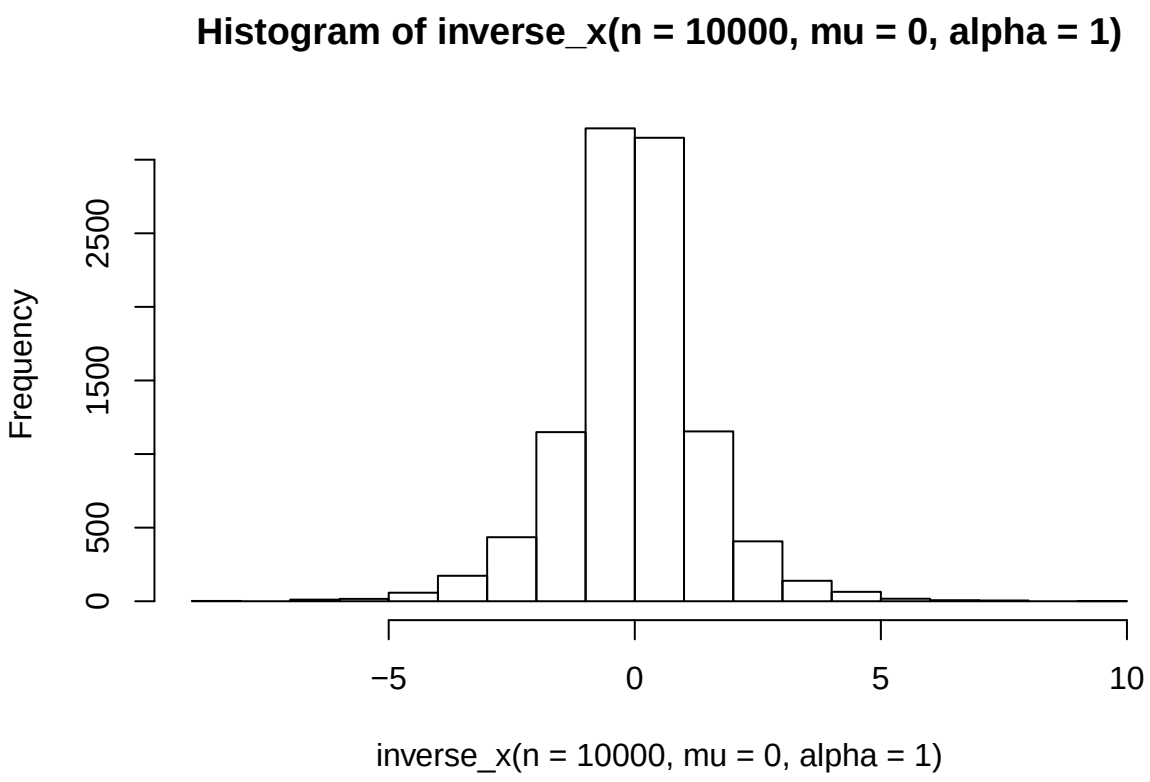
$$y = \frac{1}{2}e^{\alpha(x-\mu)}$$

$$\frac{ln(2y)}{\alpha} + \mu = x$$

$$For \ U \sim U(0,1), \qquad \frac{ln(2U)}{\alpha} + \mu = X$$

##1. Write a code generating double exponential distribution DE(0; 1) from Unif(0; 1) by using the inverse CDF method. Explain how you obtained that code step by step. Generate 10000 random numbers from this distribution, plot the histogram and comment whether the result looks reasonable.

```
inverse_x = function(n,mu,alpha){
  U <- runif(n,0,1)
  X <- ifelse(U >0.5, log(2-2*U)/(-alpha)+mu, log(2*U)/(alpha)+mu)
  return(X)
}
hist(inverse_x(n=10000,mu=0,alpha=1))
```


Histogram of inverse_x(n = 10000, mu = 0, alpha = 1)
Frequency
0
500
1500
2500
−5
0
5
10
inverse_x(n = 10000, mu = 0, alpha = 1)