Computing Stat Exercises
Khairy Mohamed
20191480897

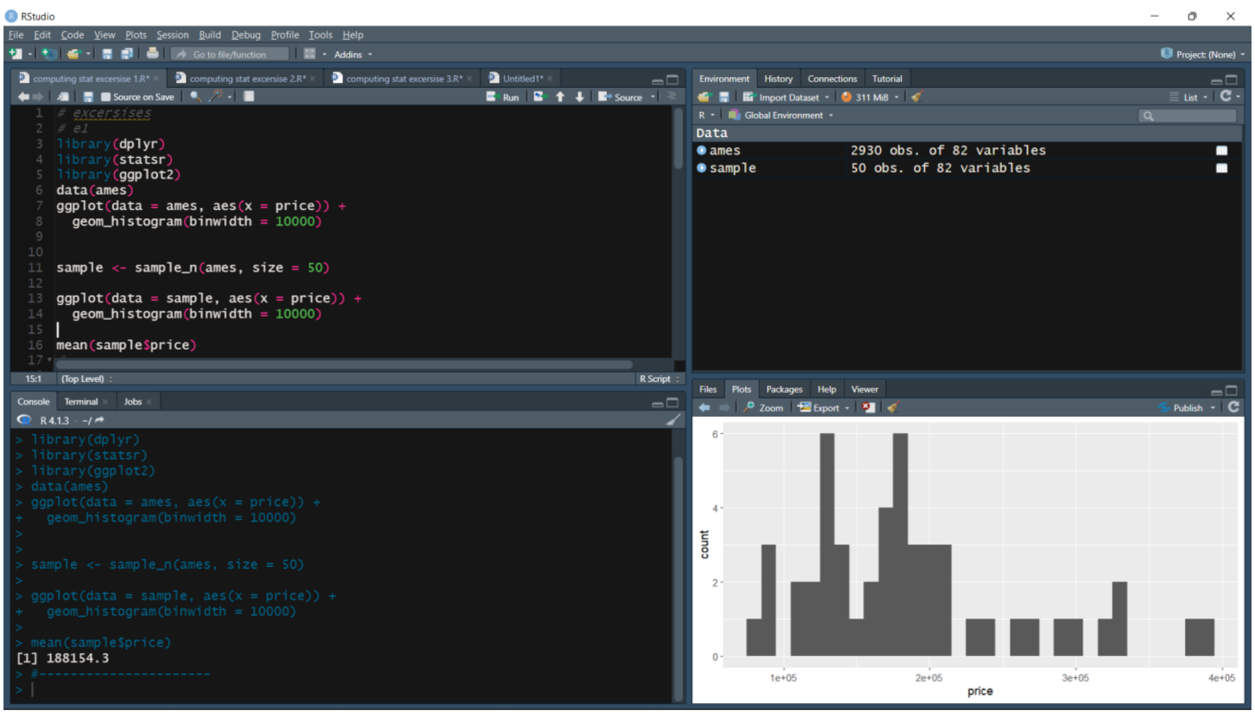
Lab 1, E1

E1: Take a random sample of size 50 from `price`. Using this sample, what is your best point estimate of the population mean?

Solution

Best point estimate of population mean with 50 samples = 188154.3

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab1, E2

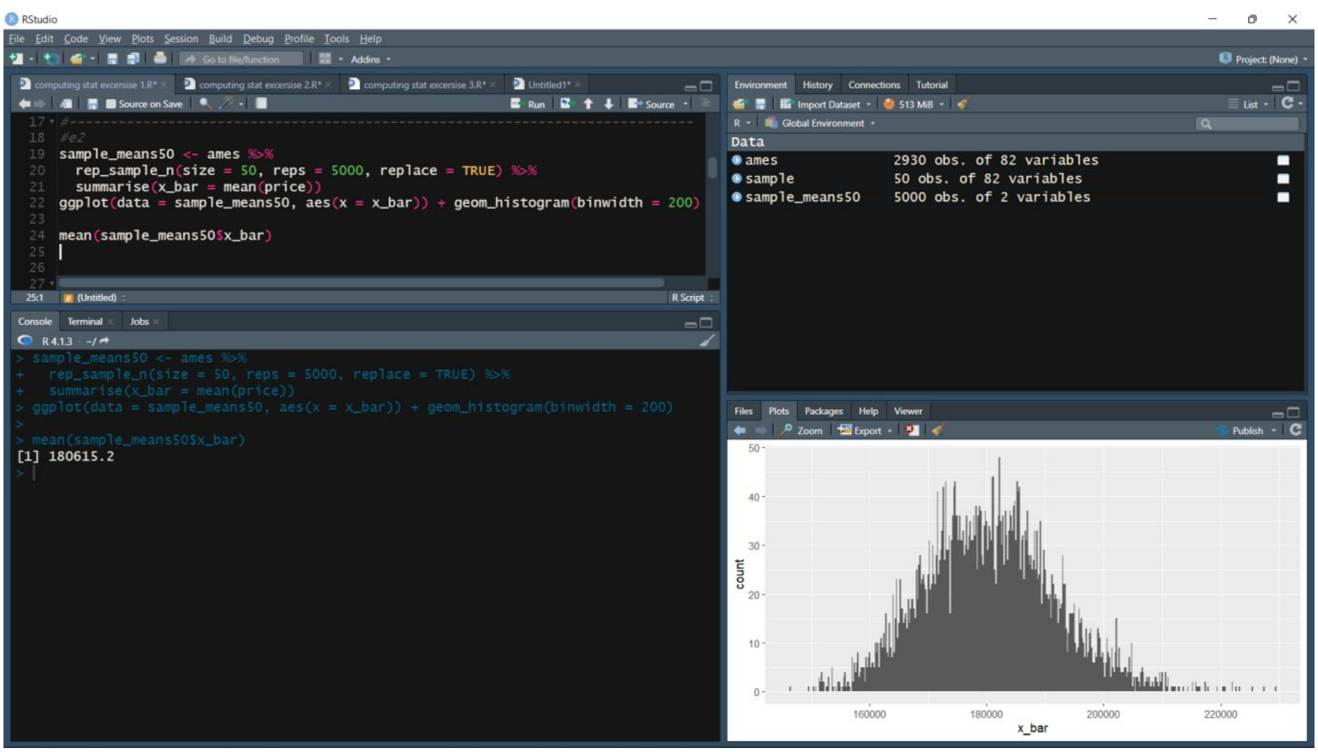
E2: Since you have access to the population, simulate the sampling distribution for $\bar{x}_{price}$ by taking 5000 samples from the population of size 50 and computing 5000 sample means. Store these means in a vector called `sample_means50`. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean home price of the population to be?

Solution

The sampling distribution is like to be normaly distributed

Estimated mean home price of population = 180615.2

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab1, E3

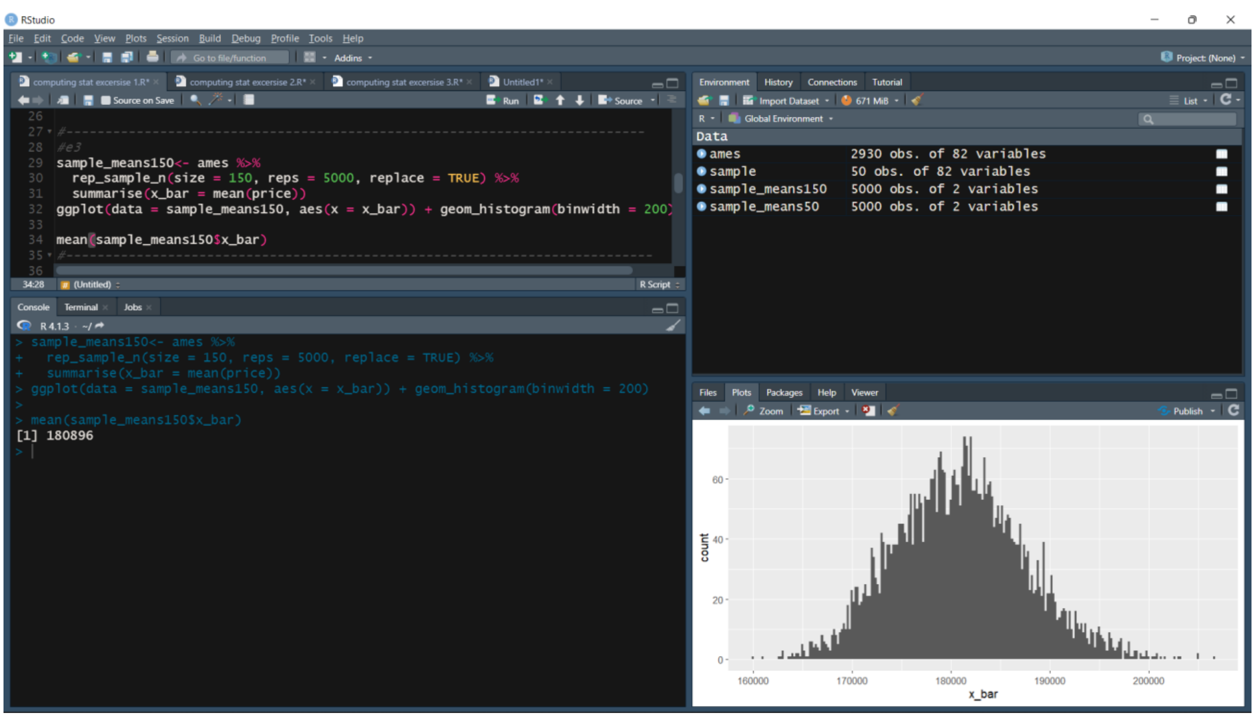
E3: Change your sample size from 50 to 150, then compute the sampling distribution using the same method as above, and store these means in a new vector called `sample_means150`. Describe the shape of this sampling distribution and compare it to the sampling distribution for a sample size of 50. Based on this sampling distribution, what would you guess to be the mean sale price of homes in Ames?

Solution

The sampling distribution is like to be normaly distributed and more perfect than that one with sample size 50

Estimated mean home price of population = 180896

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

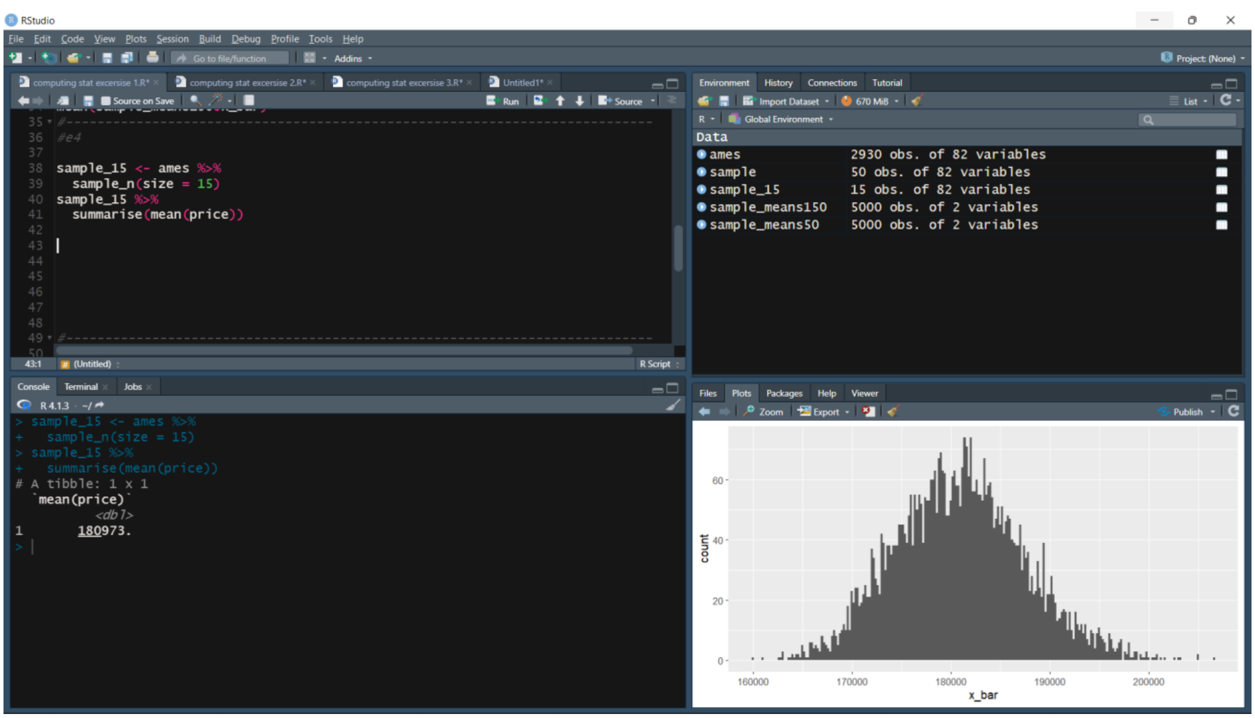
Lab1, E4

E4: Take a sample of size 15 from the population and calculate the mean `price` of the homes in this sample. Using this sample, what is your best point estimate of the population mean of prices of homes?

Solution

My best point estimate is 180973

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab1, E5

E5: Since you have access to the population, simulate the sampling distribution for $\bar{x}_{price}$ by taking 2000 samples from the population of size 15 and computing 2000 sample means. Store these means in a vector called `sample_means15`. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean home price of the population to be? Finally, calculate and report the population mean

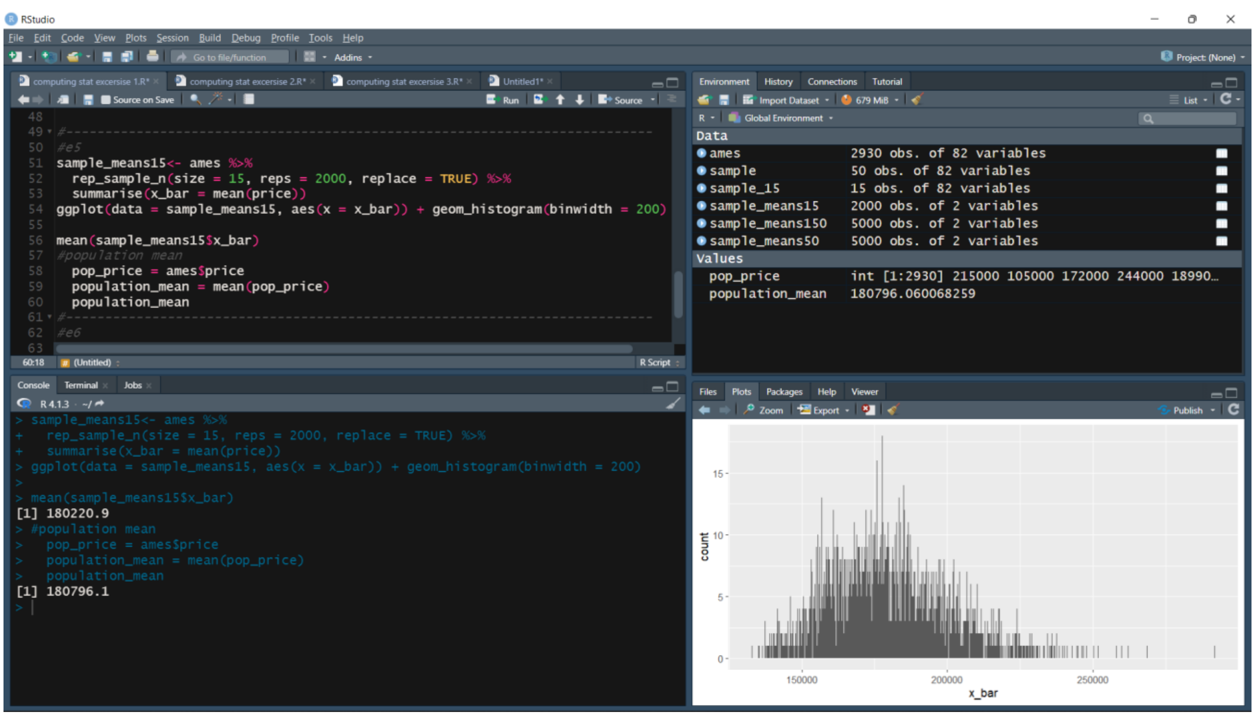
Solution

The sampling distribution is shown to be normal

Estimated population mean = 180220.9

Population mean = 180796.1

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

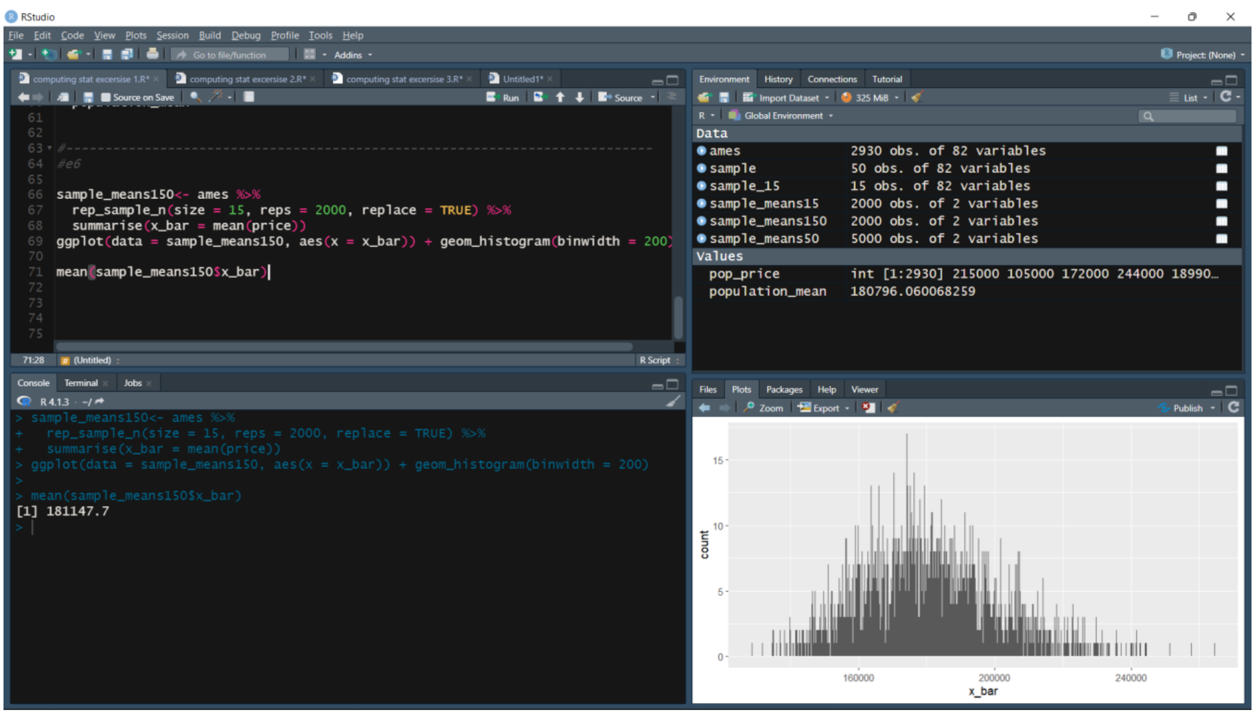
Lab1, E6

E6: Change your sample size from 15 to 150, then compute the sampling distribution using the same method as above, and store these means in a new vector called `sample_means150`. Describe the shape of this sampling distribution and compare it to the sampling distribution for a sample size of 15. Based on this sampling distribution, what would you guess to be the mean sale price of homes in Ames?

Solution

Mean of sale price = 181147.7

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab2, ECDF of BMI

- Body mass index (BMI) is a measure of body fat based on height and weight that applies to adult men and women.

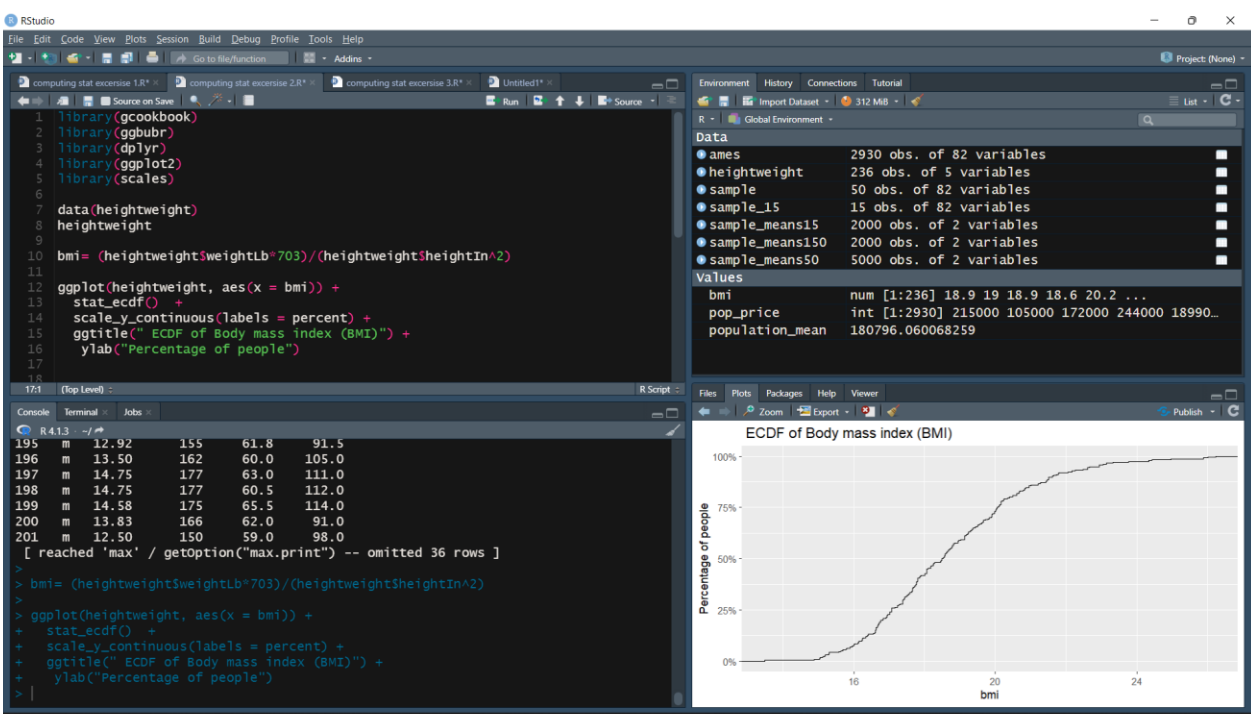
- First find the formula for the BMI and calculate it for the data set (heightweight)
- Draw the ecdf for the BMI
- According to the following BMI category and the drawn ecdf, can you determine the percentage of people that are Underweight and Overweight.

BMI Categories:
Underweight = <18.5
Normal weight = 18.5–24.9
Overweight = 25–29.9
Obesity = BMI of 30 or greater

Solution

ECDF of BMI

Code



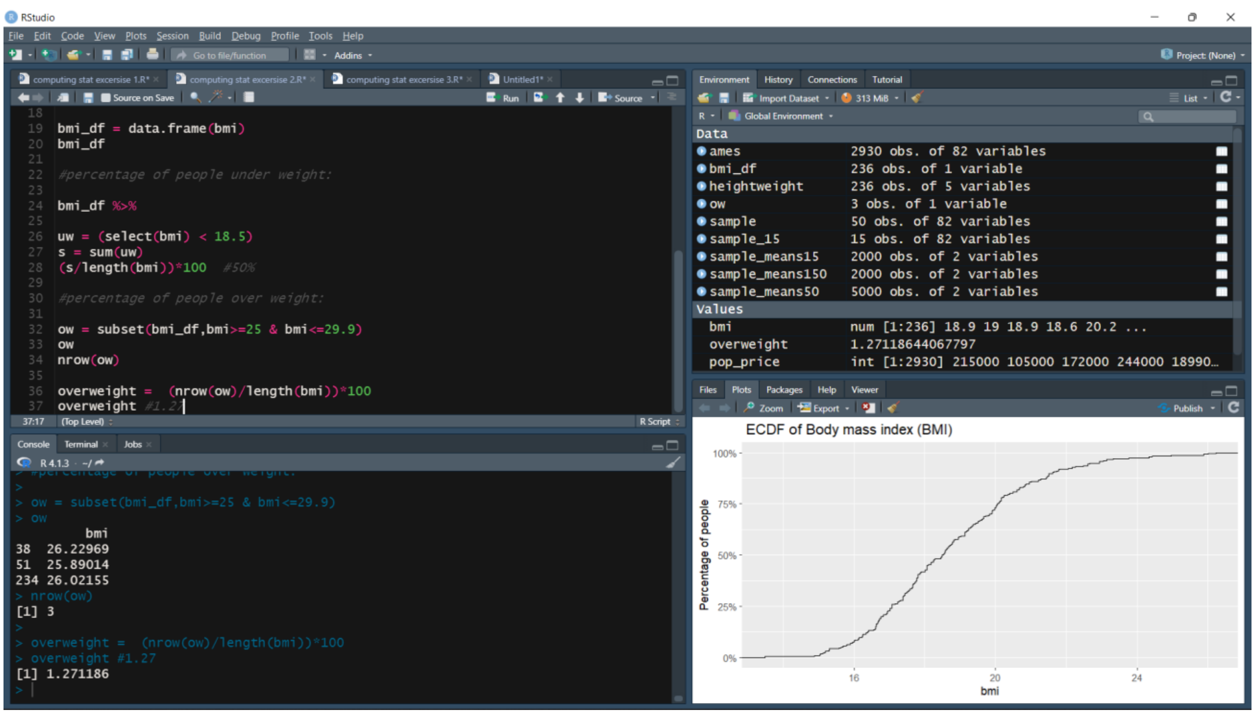
Computing Stat Exercises
Khairy Mohamed
20191480897

Lab2, Percentage of people

Solution

1) Percentage of people underweight = 50%
2) Percentage of people overweight = 1.27%

Code



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab3, E1

Exercises 1

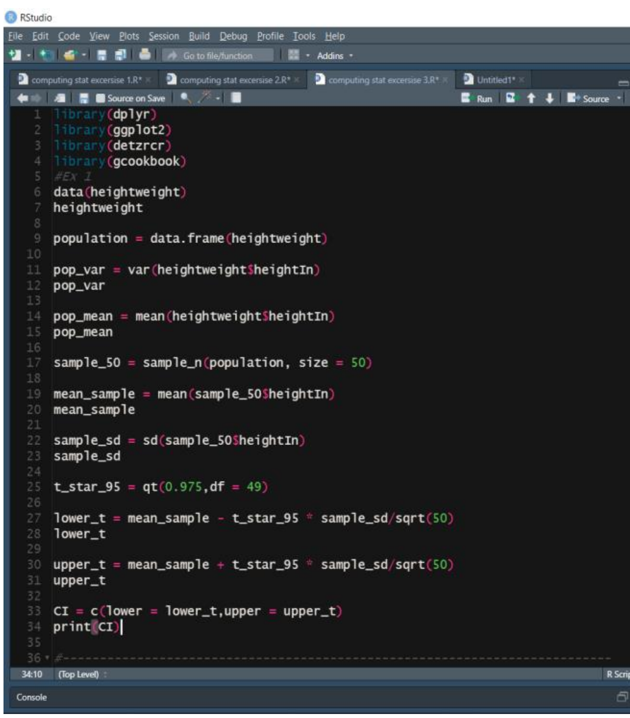
- Try to calculate the confidence interval in case of unknown population and using t-CI

Solution

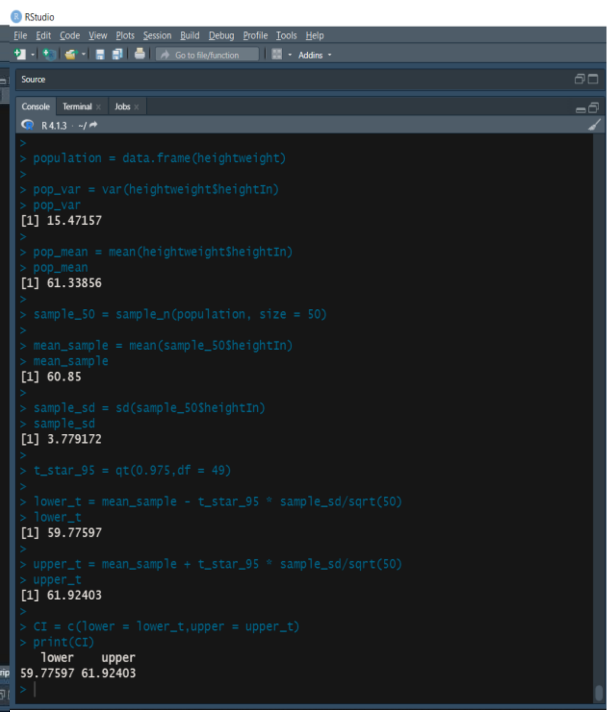
lower	upper
60.36408	62.63592

Code

Input



Output



Computing Stat Exercises
Khairy Mohamed
20191480897

Lab3, E2

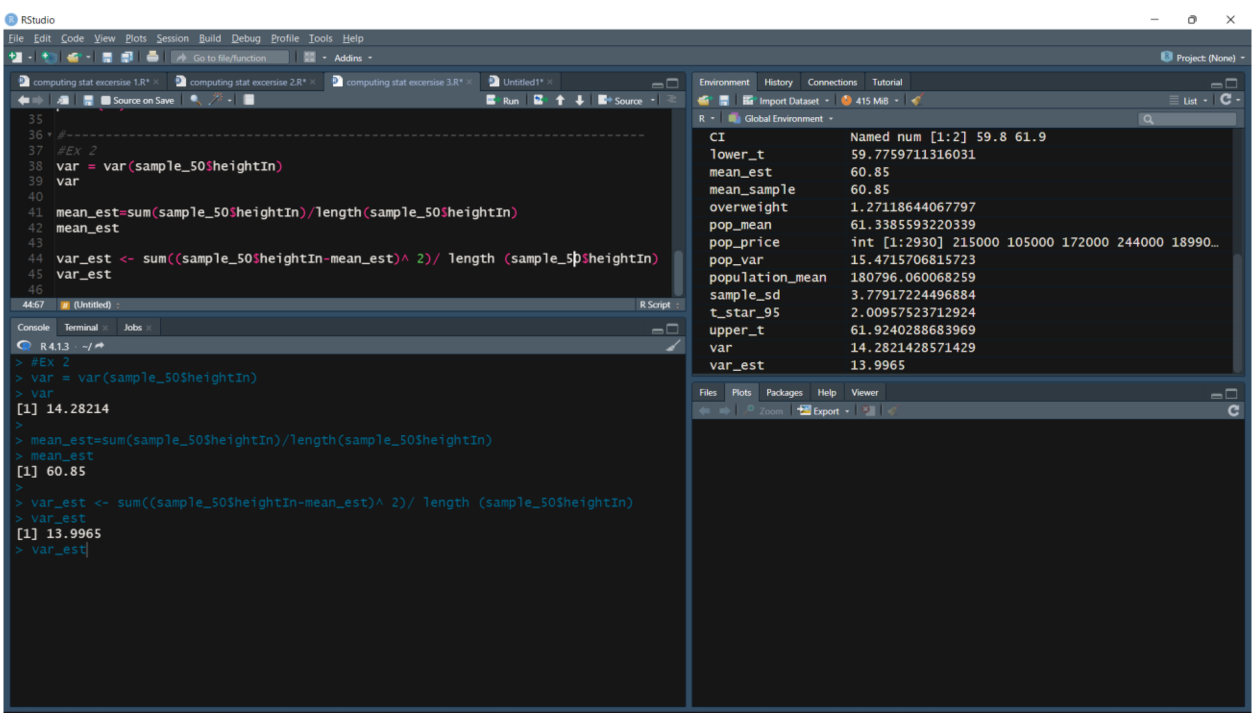
Exercises 2

1) An alternative way of obtaining the variance will be through the R function var , discover the difference between that approach and the estimated value from plug-in estimator.

Solution

The built in function var in R is more accurate than one we calculate manually because the plug-in estimator function is dependent on ECDF, so the error here is double, first when we calculate ECDF, second when we calculate plug-in variance estimator.

Code



Thank You