Practical 2

1940223, Pranav Gopalkrishna

2021-01-22

**QUESTION 1**

***Obtain an estimate of the average miles per gallon (mpg) of cars by taking a sample of size 8 using without replacement procedure.***

Sample mean is an unbiased estimator of population mean. Here, the population is all the cars in the data set, and the sample is 8 random cars chosen from the data set. The estimate of the population’s mpg will be the sample’s mpg.

# The population  
population = mtcars$mpg  
  
# To ensure we get the same random sample...  
set.seed(13)  
sampleSize = 8  
  
# Creating the sample  
sample = sample(population, sampleSize, replace = FALSE)  
sample

## [1] 13.3 22.8 18.7 19.2 17.3 18.1 10.4 15.5

# Calculating the estimated mean mpg  
est\_mean\_mpg = mean(sample)  
est\_mean\_mpg

## [1] 16.9125

**QUESTION 2**

***Check that whether the above sample using WOR provides an unbiased estimate of the population mean.***

Note that bias is the sampling error i.e. the difference between the estimate the population parameter.

# Calculating sampling error i.e. bias  
mean(population) - est\_mean\_mpg

## [1] 3.178125

# Calculating the percentage of the actual parameter the estimate reaches  
(est\_mean\_mpg/mean(population)) \* 100

## [1] 84.18105

As we can see, the bias is not 0, and is not close to 0. So, our estimate is not unbiased. However, since sample mean is an unbiased estimator in general, we can conclude that the large bias is due a small and heterogenous population, where getting representative samples is difficult. Also, our sample is small, so there is more chance that the sample is not representative.

**QUESTION 3**

***Compute the sample mean square of mpg using above sample.***

The formula for sample mean square is

(1 / (n - 1)) \* ∑(y\_i - y\_bar)^2, where

* y\_i is the ith sample element
* y\_bar is the sample mean

#  
# Defining n and y\_bar (although y\_bar = est\_mean\_mpg, we do this to avoid confusion)  
n = sampleSize  
y\_bar = mean(sample)  
  
# Finding the sum of square deviations  
sum = 0  
for(i in 1:8)  
{  
 deviation = sample[i] - y\_bar  
 sum = sum + deviation\*\*2  
}  
  
# The sum of square deviations  
sum

## [1] 102.1088

# Applying values to the formula  
sampleMeanSquare = (1 / (n - 1)) \* sum  
sampleMeanSquare

## [1] 14.58696

# Comparing the above value to the value derived from an inbuilt function  
var(sample)

## [1] 14.58696

Now we know that the inbuilt function follows the same formula. Anyways, the sample mean square of the sample is around 14.58696. So, the different mpg measures of the different cars in the sample differ on average by around 14.58696 units.

**QUESTION 4**

***Obtain standard error of an estimate of the sample taken above.***

# NOTE: Sampling is done without replacement.

The estimate i.e. the statistic we choose to find the standard error of is simply the sample mean i.e. y\_bar i.e. est\_mean\_mpg.

SE(y\_bar) = sqrt(V(y\_bar)), where

* SE(y\_bar) is the standard error of y\_bar
* V(y\_bar) is the variance of y\_bar

V(y\_bar) = ((N - n) / N) \* s^2 / n, where

* N is the population size
* n is the sample size
* s^2 is the sample mean square

#  
# Sample mean square has been calculated above, so we leave it as it is.  
# Population and sample sizes  
N = length(population)  
n = length(sample)  
  
# Variance of sample mean y\_bar  
variance = ((N - n) / N) \* sampleMeanSquare\*\*2  
  
# Standard error of sample mean y\_bar  
se = sqrt(variance)  
se

## [1] 12.63268

Hence, we may conclude that on average, the sample means taken from different samples of 8 from the given population will deviate by around 12.63268 units from their mean i.e. the mean of sample means i.e. the expected value of the sample means.

**QUESTION 5**

Obtain an unbiased estimate of population total Y i.e. N \* y\_bar.

As given in the question, it is true that N \* y\_bar is an unbiased estimator of the population total, where N is the population size and y\_bar is the sample mean (this is a corollary of the fact that sample mean is an unbiased estimator of population mean).

# Obtaining N and y\_bar  
N = length(population)  
y\_bar = mean(sample)  
  
# Obtaining the estimate of population total  
est\_popl\_total = N \* y\_bar  
est\_popl\_total

## [1] 541.2

Extra…

# Comparing it actual population total  
sum(population)

## [1] 642.9

So, according to our estimate alone, the total mpg of all the cars in the data set is 541.2. This is far from the actual sum 642.9, but that could be explained by the difficulty of drawing a representative sample from such a small and heterogenous population.