Practical 1

1940223, Pranav Gopalkrishna

2021-01-15

# OBJECTIVE

The yield in quintals for paddy crop of 50 villages in a certain area are given in the excel sheet. Select a simple random sample of size 10 units using without replacement procedure and estimate the average yield per plot along with the five number summary statistics of the population. Interpret your results.

# DATA DESCRIPTION

## The data

setwd("~/Documents/Study/statistics/statsPractical/rProgramming/rPrograms/sampling/samplingTechniques")  
myData = read.csv("data/practical1\_data.csv")  
head(myData, 10)

## yield  
## 1 20  
## 2 21  
## 3 32  
## 4 41  
## 5 55  
## 6 22  
## 7 64  
## 8 42  
## 9 28  
## 10 35

## Dimensions of the data set

dim(myData)

## [1] 50 1

The data has 50 rows i.e. entries and 1 column i.e. field.

## Certain data descriptors

summary(myData)

## yield   
## Min. :10.00   
## 1st Qu.:24.00   
## Median :29.00   
## Mean :31.82   
## 3rd Qu.:36.75   
## Max. :75.00

The data contains 50 entries, with one field that is the yield of paddy crops measured in quintals. The lowest yield measured is 10 quintals, and the highest yield measured is 75 quintals. The mean yield in this data set is 31.82 quintals.

# ANALYSIS

## Question summary

Population size: 50  
Required sample size: 10  
Estimates required: All the parameters in listed in the smmary function i.e.  
 - Minimum  
 - Maximum  
 - 1st quartile  
 - Median  
 - 3rd quartile  
 - Mean  
NOTE: Sampling must be done without replacement

sampleSize = 10  
set.seed(132)  
This function sets the seed so that the random sample generation is fixed, so the sample is random and predictable.

population = myData$yield  
sample = sample(population, 10, replace = FALSE)  
sample

## [1] 44 29 37 42 36 29 32 35 16 24

## Estimates

The estimates, compared to the parameters, are as follows

### Minimum and maximum

Sample  
min(sample)

## [1] 16

max(sample)

## [1] 44

Population  
min(population)

## [1] 10

max(population)

## [1] 75

### 1st quartile, median and 3rd quartile

Sample  
sample = sort(sample, decreasing = FALSE)  
sample[length(sample) \* 1 / 4]

## [1] 24

median(sample)

## [1] 33.5

sample[length(sample) \* 3 / 4]

## [1] 36

Population  
population = sort(population, decreasing = FALSE)  
population[length(population) \* 1 / 4]

## [1] 24

median(population)

## [1] 29

population[length(population) \* 3 / 4]

## [1] 36

### Mean

Sample  
mean(sample)

## [1] 32.4

Population  
mean(population)

## [1] 31.82

# INTERPRETATION

The variable being studied is the yeild of paddy crops measured in quintals, in a certain area. Estimations of the required parameters are as follows:

#### Minimum and maximum...

16 quintals, 44 quintals

#### 1st quartile, median and 3rd quartile...

24 quintals, 33.5 quintals, 36 quintals

#### Mean

32.4 quintals

The actual values of the parameters are as follows:

#### Minimum and maximum...

10 quintals, 75 quintals

#### 1st quartile, median and 3rd quartile...

24 quintals, 29 quintals, 36 quintals

#### Mean

31.82 quintals

As we can see, except the minimum and maximum values, the estimates derived from the sample are very close to the population. Hence, we may consider the sample as representative of the population, since minimum and maximum values are anyways very subject to change when measured across random sample, since it requires the selection of a specific set of values at the extremities of the population.