Stat 123 Homework Assignment 1

Koki Itagaki

2023-01-26

#Q1 Twenty bottles of water were randomly selected from a large collection of bottles in a company's warehouse. These twenty bottles are referred to as the:

a) Parameter
b) Population
c) Sample

Answer) c

#Q2 A journalist wanted to know how his constituents felt about a planned rezoning. He randomly selected 234 names from the city phone directory and conducted a phone survey.

##a) What is the population of interest? ##b) What is the sample of interest? ##c) What type of sampling was used?

Answer: a)all constituents in the city

b)234 names from the city phone directory

c)simple random sample

#Q3 You can see a sample of dataset below. #Patient ID Cancer Type Number of Visits Status #123_87 Lung 3 Death #134_99 Brain 4 Survival #135_46 Breast 5 Death #124_76 Kidney 3 Death #132_54 Lung 6 Survival

(a) What are the individuals in the data set?
(b) What are the variables in the data set?
(c) Determine if each variable is categorical or numerical

Answer: (a)Patient ID

(b)Cancer type, Number of visits, and Status

(c)Cancer type: Categorical variables Number of visits: Numerical variables Status: Categorical variables

#Q4. A journalist is trying to determine the average age of the BC residents who have received fines #for violating restrictions related to COVID-19. He solicits data from a local police department #and records the following ages for the five tickets given out by this precinct. #13; 10; 25; 34; 15

```
Sample<- c(13,10,25,34,15)
```

```r
#(a) What is the parameter of interest?
print("My answer : The average age of the BC residents who have received
fines for violating restrictions related to COVID-19")
```

```
## [1] "My answer : The average age of the BC residents who have received
fines for violating restrictions related to COVID-19"
```

```r
#(b) Estimate the sample varience. Round your answer to 2 decimal places.
sample_vari<- round(var(Sample),2)
sample_vari
```

```
## [1] 98.3
```

```r
#(c) Estimate the sample mean. Round your answer to 2 decimal places.
x <- mean(Sample)
round_x<-round(x,2)
round_x
```

```
## [1] 19.4
```

(a) The average age of the BC residents who have received fines for violating restrictions related to COVID-19

#Q5 Create the following data frame called Sample using the data.frame() function. ##ID Name Age Vote ##1 Juan 22 TRUE ##2 Maria 15 FALSE ##3 Mark 19 TRUE

```r
Sample<- data.frame(ID = c(1,2,3), Name = c("Juan", "Maria", "Mark"), Age =
c(22,15,19), Vote = c(TRUE,FALSE,TRUE))
Sample
```

```
##   ID  Name Age  Vote
## 1  1  Juan  22  TRUE
## 2  2 Maria  15 FALSE
## 3  3  Mark  19  TRUE
```

```r
#(a) Create a character vector called Age which contains the values from the
second column of the Sample.
Age <- Sample[,2]
Age
```

```
## [1] "Juan"  "Maria" "Mark"
```

```r
#(c) Calculate the number of TRUE votes in the third column.

NumT<- sum(Sample$Vote == TRUE)
NumT
```

```
## [1] 2
```

```r
#(d) Calculate the average of the ages in the second column.
Average <- mean(Age)
```

```
## Warning in mean.default(Age): argument is not numeric or logical:
returning NA

Average

## [1] NA
```

#Q4. Create a list called CourseName and mention the name of courses you have this semester and then create the second list called CourseUnits with the units of each course.

```
CourseName <- list("Stat 123", "Stat 354", "Stat359", "Seng265")
CourseName

## [[1]]
## [1] "Stat 123"
##
## [[2]]
## [1] "Stat 354"
##
## [[3]]
## [1] "Stat359"
##
## [[4]]
## [1] "Seng265"

CourseUnits <- list(1.5,1.5,1.5,1.5)

#(a) Combine the list CourseName and CourseUnits together.
list<- c(CourseName,CourseUnits)
list

## [[1]]
## [1] "Stat 123"
##
## [[2]]
## [1] "Stat 354"
##
## [[3]]
## [1] "Stat359"
##
## [[4]]
## [1] "Seng265"
##
## [[5]]
## [1] 1.5
##
## [[6]]
## [1] 1.5
##
## [[7]]
## [1] 1.5
```

```
##
## [[8]]
## [1] 1.5

#(b) Change the values of CourseUnits to the character.
CourseUnits<-as.character(CourseUnits)

class(CourseUnits)

## [1] "character"
```