HomeworkAssigment#1

Eric Huber

01/02/2022

#Question 1: 1. A farmer wants to determine the proportion of carrot seeds planted in her field that successfully grow into carrots. It would take too much time to count the total amount of seeds planted in the field and the total yield of carrots that result. Thus, she decides that she needs to take a sample to estimate this proportion.

Part A: State the population and the variable of interest to this farmer.

(1.a) The population would be all of the carrot seeds planted in her field. The variable is the amount of seeds that successfully reach adulthood and grow into carrots.

Part B: Give an example of a way the farmer could perform a convenience sample.

(1.b) The farmer could test the row of carrots planted closest to where she lives.

Part C: Give an example of a way the farmer could perform a simple random sample.

(1.c) The farmer could split her land into a grid and name or number each square of the grid. Then she could write each one of the grid numbers on to pieces of paper, put them into a hat and pick a number of them relevant to the number of grids named.

Part D: Give an example of a way the farmer could perform a stratified random sample.

(1.d) The farmer could divide the farm land into different groups based on how much sun the get per day. She could then conduct a simple random sample on those groups in a method similar to the one highlighted above.

Part E: What is the population proportion of interest? What would be a good statistic to use to estimate the population parameter?

(1.e) The population proportion of interest is the value of the percentage of the carrot seeds that are successfully grow into carrots. What I mean by that is if 74.3% of carrots reach full maturity, 743 would be the population proportion. A statistic you could use would be X/N, where N is the size of your sample and X is the count of successfully grown carrots.

Question 2:

2. The following question deals with the data set lynx which is one of the built-in data sets included with R.

Part A: Describe what information is contained in the data set. How did you determine this?

(2.a) The data set lynx contains the annual numbers of lynx trappings in Canada from 1821-1934. First I checked to see if it was an available data set by typing data() and then when I found the data set I typed ?lynx which opens a help window with information about the data set. The code I use is put below

```
data()
?lynx

## starting httpd help server ... done
```

Part B: Create a character vector called 'years' which contains the years of the trappings.

```
# 2.b
years = c(1821:1934)
```

Part C: Set the names of the lynx vector equal to years.

```
# 2.c
data("lynx")
names(lynx)=years
```

Part D: How many lynx were trapped in 1899?

(2.d) 153 lynx were trapped in 1899

Part E: What is the average number of lynx trappings in a year?

```
mean(lynx)

## [1] 1538.018
```

(2.e) The average number of lynx trappings is 1538.018

Question 3

3. The following question deals with the data set vegas which can be found in Brightspace by clicking on Content – > Homework Assignments. This data set represents the winnings and losses of a group of friends who went to Las Vegas together.

Part A: Use the head() function to determine the games these friends played in Vegas.

```
vegas = read.csv("vegas.csv")
head(vegas)
```

```
##     Name BlackJack  Poker   Slots Roulette   Craps
## 1  Betty     50.46  41.68  262.88  -114.46  106.59
## 2   John      6.80   4.00  212.70    48.46  890.84
## 3 Dwayne    -98.29 -54.82  252.58   -66.82   38.65
## 4 Sophia    183.73  59.49   95.19  -115.82   15.20
## 5  Luisa     43.12  38.79  -10.95  -230.82   29.88
## 6 Carlos     49.40  68.40 -289.88    53.92 -275.19
```

(3.a) My friends Betty, John, Dwayne, Sophia, Luisa and Carlos are the friends I played games with.

Part B: Create a character vector called friends which contains the values from the first column of the data set.

```
friends = vegas$Name
```

Part C: Using the R command as.matrix(), create a matrix called gameResults which contains all the columns except the first one from the vegas data set

```
gameResults = as.matrix(vegas[ ,2:6])
gameResults
```

```
##       BlackJack  Poker   Slots Roulette   Craps
##  [1,]     50.46  41.68  262.88  -114.46  106.59
##  [2,]      6.80   4.00  212.70    48.46  890.84
##  [3,]    -98.29 -54.82  252.58   -66.82   38.65
##  [4,]    183.73  59.49   95.19  -115.82   15.20
##  [5,]     43.12  38.79  -10.95  -230.82   29.88
##  [6,]     49.40  68.40 -289.88    53.92 -275.19
##  [7,]    343.51 -59.21   74.33    85.42 -577.07
##  [8,]    284.26  24.08  -67.36    72.55 -370.38
##  [9,]   -233.37   7.29  -76.96    84.13 -519.10
## [10,]    -81.39  42.39   -6.81   -59.80 -110.55
```

Part D: Create a vector called totals which contains the row sums of the matrix gameResults. What do the values in this vector represent?

```
totals = c(rowSums(gameResults))
totals
```

```
##  [1]  347.15 1162.80   71.30  237.79 -129.98 -393.35 -133.02  -56.85 -
738.01
## [10] -216.16
```

Part E: Set the names of the vector totals equal to friends.

```
names(totals) = friends
```

Part F: Use the R functions min() and max() to determine which friend won the most money and which friend lost the most money in Vegas.

```
min(totals)

## [1] -738.01

max(totals)

## [1] 1162.8
```

(3.f) The friend who lost the most was Calum and my friend who lost the most was John!

Part G: What was the average amount of money won or lost by the group of friends on the trip?

```
mean(totals)

## [1] 15.167
```

(3.g) The average amount of money won was $15.167! A shockingly successfully vegas trip!

Part H: Set a seed to 34. Take a random sample of 4 of the friends and determine the sample average amount of money won or lost by the sample of friends on the trip.

```
set.seed(34)

Samp = sample(totals, 4)
mean(Samp)

## [1] 178.7725
```