Homework Assignment 1

Parker DeBruyne - V00837207

2022-01-27

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.

(a) State the population and the variable of interest to this farmer.

Population: All carrot seeds.
Variable: Whether or not they grew into carrots.

(b) Give an example of a way the farmer could perform a convenience sample.

She could take a sample from the closest patch to her house.

(c) Give an example of a way the farmer could perform a simple random sample.

She could assign each seed intended to be planted with digets n (equal to the number of seeds), create a data table of random numbers, select a random row, and then scan until one of the assigned numbers for the objects is found. Repeat until the desired sample size has been chosen.

Or draw a random amount out of a bag.

(d) Give an example of a way the farmer could perform a stratified random sample.

If the farmer believed there would be varience between groups, she could divide her field into different section (such as closest to a water feature, trees, rocks, or sunny patches) and then take a SRS from each group.

(e) What is the population proportion of interest? What would be a good statistic to use to estimate the population parameter?

The proportion of seeds that successfully grow, to the seeds that do not.

The Population Proportion statistic p, which is the number of individuals with a certain characteristic divided my the number of individuals in the population.

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

(a) Describe what information is contained in the data set. How did you determine this?

This data set contains information on the annual numbers of lynx trappings for 1821-1934 in Canada. Taken from Brockwell & Davis (1991), this appears to be the series considered by Campbell & Walker (1977).

I obtained this description by first loading the data using the data command and then by quering the data set.

```
data(lynx)
?lynx
```

(b) Create a character vector called years which contains the years of the trappings.

```
years = c(1821:1934)
```

(c) Set the names of the lynx vector equal to years.

```
names(lynx) = years
```

(d) How many lynx were trapped in 1899?

```
lynx["1899"]
```

```
## 1899
##  153
```

153

(e) What is the average number of lynx trappings in a year?

```
mean(lynx)
```

```
## [1] 1538.018
```

1538.018

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.

(a) Use the head() function to determine the games these friends played in Vegas.

```
setwd("/Users/admin/Documents/School/STAT 123/homework/homework1")
getwd()
```

```
## [1] "/Users/admin/Documents/School/STAT 123/homework/homework1"
```

```
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
```

```
class(vegas)
```

```
## [1] "data.frame"
```

```
games = names(vegas)
games
```

```
## [1] "Name"      "BlackJack" "Poker"     "Slots"     "Roulette"  "Craps"
```

(b) Create a character vector called friends which contains the values from the first column of the data set.

```
friends = vegas$Name
friends
```

```
##  [1] "Betty"     "John"      "Dwayne"    "Sophia"    "Luisa"     "Carlos"
##  [7] "Andrew"    "Charlotte" "Calum"     "Layla"
```

(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.

```
length(vegas)
```

```
## [1] 6
```

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

(d) Create a vector called totals which contains the row sums of the matrix gameResults. What do the values in this vector represent?

```
totals = 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
```

These values represent the total winnings/losses of each friend across all games.

(e) Set the names of the vector totals equal to friends.

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

```
##     Betty      John    Dwayne    Sophia     Luisa    Carlos    Andrew Charlotte
##    347.15   1162.80     71.30    237.79   -129.98   -393.35   -133.02    -56.85
##     Calum     Layla
##   -738.01   -216.16
```

(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.

```
findmin <- totals == min(totals)
findmin
```

```
##     Betty      John    Dwayne    Sophia     Luisa    Carlos    Andrew Charlotte
##     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
##     Calum     Layla
##      TRUE     FALSE
```

```
totals[findmin]
```

```
##   Calum
## -738.01
```

```
findmax <- totals == max(totals)
findmax
```

```
##     Betty      John    Dwayne    Sophia     Luisa    Carlos    Andrew Charlotte
##     FALSE      TRUE     FALSE     FALSE     FALSE     FALSE     FALSE     FALSE
##     Calum     Layla
##     FALSE     FALSE
```

```
totals[findmax]
```

```
##   John
## 1162.8
```

Calum lost the most money, and John won the most money.

(g) What was the average amount of money won or lost by the group of friends on the trip?

```
mean(totals)
```

```
## [1] 15.167
```

$15.167

(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)
mean(sample(totals, 4, replace=F))
```

```
## [1] 178.7725
```

$178.7725