Assignment 1

Branden Walter

2023-05-26

Question 1

A farmer wants to determine the proportion of pumpkin seeds planted in her field that successfully grow into pumpkins. It would take too much time to count the total number of seeds planted in the field and the total yield of pumpkins. 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 = Seeds planted by the farmer.

Variable = Whether a planted pumpkin seed grows.

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

The Farmer manually counting the pumpkins as he walks.

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

He could close his eyes, spin around a few times and then point at a random pumpkin, repeat until sample aquired.

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

If he sections the field into sections and then takes a random sample from each of the sections

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

Number of successful pumpkins grown would be a good parameter.

Question 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. What command in R did you use to find this information?

```
#help(lynx) # Commented it out since it kept starting a webserver with the info....
```

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

```
years <- as.character(time(lynx))
```

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

```
names(lynx) <- years
```

(d) How many lynx were trapped in 1901? Use years as your index.

```
lynx["1901"]
```

```
## 1901
##  758
```

(e) What is the average number of annual lynx trappings from 1821 to 1920, inclusive? Hint: You want to compute the average trappings for the first 100 data points.

```
mean(lynx[1:100])
```

```
## [1] 1527.77
```

Question 3

> The following question deals with the data set casino 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 a local casino together.

(a) Read in the data file and name the data frame casino.

```
casino <- read.csv("casino.csv")
```

(b) Use the head() function to determine the games these friends played in the casino.

```
head(casino)
```

```
##     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
```

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

```
friends <- as.character(casino$Name)
```

(d) Using the R command as.matrix(), create a matrix called winnings which contains all the columns except the first one from the casino data set.

```
winnings <- as.matrix(casino[,-1])
winnings
```

```
##       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
```

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

> The values in the `totals` vector represent the total money won or lost by each friend.

```
totals <- rowSums(winnings)
totals
```

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

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

```
names(totals) <- friends
```

(g) Use the R functions min(), max(), which.max() and which.min() to determine which friend won the most money and which friend lost the most money in the casino.

```
friends_won <- names(totals)[which.max(totals)]
friends_lost <- names(totals)[which.min(totals)]
friends_lost
```

```
## [1] "Calum"
```

```
friends_won
```

```
## [1] "John"
```

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

```
average_winning <- mean(totals)
average_winning
```

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