

Homework 1 Solutions

Question 1:

- (a) The population is all pumpkin seeds and the variable of interest is whether or not the seed grows into a pumpkin (you can think of this as a logical variable which takes on values TRUE or FALSE).

(Note: there are many correct answers to part b, c and d)

- (b) A convenience sample could involve the farmer observing the seeds that are planted closest to her house.
- (c) The farmer could put all the seeds in a bag and then randomly select a sample. Then she could mark where each of those seeds get planted and observe whether or not they grow into pumpkins.
- (d) The farmer could section her field into pieces (perhaps based on quality of soil or amount of sunlight that it gets) and then observe a random sample of seeds from each part of the field.
- (e) The population parameter of interest is the proportion of pumpkin seeds that grow into pumpkins. A good statistic to estimate this parameter would be the sample proportion \hat{p} .

Question 2:

```
# (a) Describe the data in the lynx data set.
```

```
?lynx
```

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

```
# Using the help function in R, we can pull up information on the lynx data  
# set. This gives a description of the data which contains the annual  
# numbers of lynx trappings for 1821-1934 in Canada.
```

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

```
years = as.character(c(1821:1934))
```

```
# years = c(1821:1934) will create a numerical vector.
```

```
# (c) Set the names of the lynx vector to years.  
names(lynx) = years
```

```
# (d) How many lynx were trapped in 1901?  
lynx["1901"]
```

```
## 1901
## 758
# We see that there were 758 lynx trappings in 1901.

# (e) What is the average number of lynx trappings from 1821 to 1920 inclusive?
mean(lynx[1:100])

## [1] 1527.77
# The average number of lynx trappings from 1821 to 1920 was 1527.77.
```

Question 3:

```
# (a) Read in the data set.
casino = read.csv("casino.csv")

# (b) Use the head() function to determine the games the friends played.
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

# The games are BlackJack, Poker, Slots, Roulette, and Craps.

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

# you could also do this by using the index of the first column:
# friends = casino[,1]

# (d) Create a matrix called winnings which contains all the columns
# except the first one from the casino data set.
winnings = as.matrix(casino[,2:6])

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

# The sum of the rows represents how much money each person won or lost
# while in the casino.

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

# (g) Determine which friend won the most money and which friend lost the
# most money in the casino.
totals[which.max(totals)]
```

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

We see that John won the most money in the casino winning \$1162.80

```
totals[which.min(totals)]
```

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

We see that Column lost the most money in the casino losing \$738.01

(h) What was the average amount of money won or lost by the group

```
mean(totals)
```

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

On average the friends won a total of \$15.17