Stat 21 Week 2 Worksheet

Code ▾

Person 1, Person 2, Person 3, etc.

Wednesday, Jan 26, 2022

This worksheet is to be completed in class as a group. Take a minute to have everyone introduce themselves and then choose who will play the role of recorder and reporter. The recorder is responsible for completing this worksheet and emailing the .Rmd file to your professor with the subject "Stat 21: Week 2 Worksheet". Make sure everyone's names are included at the top of this document! The reporter is responsible for taking note of your answers and any questions that come up to share them with Prof. Suzy when she stops by to check in on your group. Other groups members are responsible for keeping the discussion on-task and providing possible solutions or sharing questions with the group.

For this assignment, I suggest that you choose the recorder to be whoever went to sleep the earliest last night and choose the reporter to be whoever went to sleep the latest.

HW 1 Problem 3

Below is a stem-and-leaf plot for the profits (as percent of sales) for 29 different corporations in the US. The stems are split so that each stem represents a span of 5%. Thus the smallest observation is a loss of 9% and the largest observation is a gain of 25%. As another example, -0|3 is interpreted as a a loss of 3%.

```
-0|9 9
-0|1 2 3 4
 0|1 1 1 1 2 3 4 4 4
 0|5 5 5 5 6 7 9
 1|0 0 1 1 3
 1|
 2|2
 2|5
```

Instructions

Referring to Problem 3 from HW 1, answer the following questions with your group mates. Note, none of these questions require any R coding, instead have the group recorder write your groups answers in the space provided below.

Question 1: How would you describe the distribution of the data from this stem-and-leaf plot?

Answer:

Question 2: How could you calculate the median value of the data without using R?

Answer:

Question 3: What is the formula to calculate the standard deviation of this numeric sample? (Just write out, no need to perform the calculations.)

Answer:

This HW 1 problem asks you to calculate some common *summary statistics* for a particular data set. These summary statistics are so commonly used that we collecitvely refer to them as a *six-number summary*. The six-number summary of a quantiative data set includes

- the mean and median,

- lower and upper quartiles (that is, the lower 25th quantile and lower 75th quantile),

- minimum and maximum (which together describe the *range*).

Other useful summary statistics regarding the spread of the data are the *variance* and *standard deviation*. Calculating these summary statistics by hand is not necessary and here you will practice calculating them with R code.

The R code chunk below imports a vector of quantitative data points (vector just means there's more than one) and then uses R's default functions to calculate the summary statistics mentioned above:

Hide

```
vector_of_data <- c(1,1,2,1,3,1,4,-1,-2)

mean(vector_of_data)
```

```
## [1] 1.111111
```

Hide

```
sd(vector_of_data)
```

```
## [1] 1.833333
```

Hide

```
var(vector_of_data)
```

```
## [1] 3.361111
```

Hide

```
summary(vector_of_data)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.000   1.000   1.000   1.111   2.000   4.000
```

Instructions continued

The R code chunk below reads the data from the stem-and-leaf plot in Problem 3 of HW 1 into our R workspace and gives this data set the name `stem_data`:

```
stem_data <- c(-0.09, -0.09, -0.01, -0.02, -0.03, -0.04,
               0.01, 0.01, 0.01, 0.01, 0.02, 0.03, 0.04, 0.04, 0.04,
               0.05, 0.05, 0.05, 0.05, 0.06, 0.07, 0.09,
               0.10, 0.10, 0.11, 0.11, 0.13, 1.00, 0.22, 0.25)
```

Work with your group to complete the R code chunk below to calculate the summary statistics mentioned above for the `stem_data`.

```
## Erase this comment and write your code here
```