

Lab 2

Kristina Arevalo

Wed Sep 09 2020

Contents

Problem 1	1
Problem 2	1
Problem 3	2
Problem 4	2
Problem 5	2
Problem 6	2

Problem 1

Use R to demonstrate that the mean minimizes the sum of the squared deviations from the mean. Accomplish the following steps:

- Produce a sample of at least 10 or more different numbers
- Produce a simulation following the example from the concepts section
- Use your simulation to test a range of numbers smaller and larger than the mean to show that the mean minimizes the sum of the squared deviations from the mean.
- Plot your results.

```
a <- 1:15  
my_sample <- sample(a, replace = TRUE)
```

Problem 2

Write a custom R function for any one of the following descriptive statistics: median, mode, standard deviation, variance. Demonstrate that it produces the same value as the base R function given some set of numbers.

Problem 3

Imagine the same instructor taught a morning, afternoon, and evening section of the same course. And, the average scores for each section on a midterm were 85% for the morning, 90% for the afternoon, and 93% for the evening sections. Create a `data.frame` representing these means for each section. Then, use `ggplot2` to plot the means as bar graph. (hint you will need one vector for the class sections, and one vector for the means. Then you can combine them into a `data.frame` before plotting them)

Problem 4

Imagine there were two instructors, and they each taught different sections in the morning, afternoon and evening. The midterm averages for instructor 1 were 75%, 78%, and 80% in the morning, afternoon, and evening. The midterm averages for instructor 2 were 88%, 76%, and 63% for the morning, afternoon, and evening. Create a `data.frame` representing the means, the time of day, and the instructors (three columns). Then plot `data.frame` using `ggplot2` as a bar graph.

Problem 5

Import the `WHR2018.csv` data file, containing measure from the World Happiness report from 2018. For the years 2010 to 2015, what was the mean “healthy life expectancy at birth” for each year (find the mean for each year across countries). Show your results in a table and in a graph using `ggplot`.

Problem 6

Repeat the above, except in addition to calculating the mean for each year, also calculate the standard deviation for “healthy life expectancy at birth” for each year. Then, add error bars to your graph using the `+1` and `-1` standard deviations from the means for each year.