

Advanced Psychological Statistics
PSYCH-UA.11
Department of Psychology
Spring 2022

Data Assignment 1

1. A researcher collects the years of education for a sample group of participants. The ages are

Participants	Years of Education
1	28
2	23
3	28
4	30
5	24
6	30
7	20
8	25
9	29
10	24
11	24
12	24
13	20
14	28
15	29

1. Create a data set in R using the above data

What is the mean?

What is the median?

What is the mode?

What can you say about the three measures of central tendency?

2. Create a random selection of 10 numbers ranging from 1 to 20 (use sample.int)

What is the median?

What is the mean? Is there a difference between the two? Why?

3. A researcher wants to study the impact of the presence (in the room) of a sweet snack on task-completion. 30 participants are given 5 logic problems to solve. Half of the participants are randomly assigned to desks that have only a pencil and the word-problems. The remaining participants are assigned to desks with a pencil, the same word-problems and a candy dispenser. Participants are timed and the completion times are recorded.

These are the times (in seconds)

Candy-Absent Group			Candy-Present Group	
Participant	Reaction Time		Participant	Reaction Time
1	501		16	690
2	536		17	691
3	659		18	510
4	317		19	586
5	530		20	675
6	523		21	470
7	381		22	533
8	573		23	693
9	535		24	440
10	509		25	614
11	604		26	475
12	704		27	374
13	370		28	500
14	440		29	478
15	404		30	664

Compute the mean and median for both group

Candy-Absent

Candy-Present

Mean

Median

What do you think about the results you've computed?

If you changed the highest score in the Candy-Absent group to be 10 times the original value, what would happen to the mean? What about median?

If you changed the highest score in the Candy-Present group to be one tenth its original value, what would happen to the mean? What about the median?

4. Create a set of 50 numbers using the range 1 to 5. (Taking 50 numbers from the set (1,2,3,4,5) requires that each “draw” is put back into the original set.) Show your plot. What is the mode of your data set?
5. Create a set of 10 numbers ranging from 1 to 100. (The sample.sample.int would be helpful here as well.) Describe the output. Identify the approximate value of the median using the plot. Compare your visual estimate to the actual value (use the “median” command). What can you conclude about the dispersion of the values of the data set based on the plot?

Advanced topic (...if you're interested)

Create a normal distribution

Create a data set in any range using the “seq” command. For a range of numbers, this creates a sequence from the first, to the second number, by step value you select. Try -

```
> x = seq(-5, 5, by = .1)
```

Next find the probability values, based on the normal distribution, that is associated with each value in x. (The values in ‘x’ will appear on the x-axis). To find what probability value, based on the normal distribution, is associated with each value value in x use the “dnorm” command. The “dnorm” command returns a probability (again from the normal distribution) associated with the values of the x-axis, and based on where the mean is. (Since our range is -5 to 5, an intuitive mean could be 0.)

```
> y = dnorm(x, mean = 0, sd = 1.5)
```

Now, plot x and y. The plot shows the probability (along the y-axis. The higher the height, the greater the likelihood) of each value (on the x-axis) of the original data set.

```
> plot(x,y)
```

Note – the default plot will display each x,y pair as a circle. Often lines are used. To change the circles to a line, change the “type” of the plot

```
> plot(x, y, type = "l") (that's an “el”, not a one)
```

Find the three measures of central tendency? What do you notice about these? How can you account for similarities and differences?