Measures of Central Tendency

In addition to analyzing numerical data by dividing it into class intervals and exploring frequencies for each interval, you can also make certain statements about a set of data as a whole. In this section, we will discuss the measures of central tendency – the mean and the median. They describe the "average" or "central" observation around which the data tends to cluster or concentrate.

Mean

The mean is commonly used in our daily lives. It is the arithmetic average that is calculated by adding up all observations and dividing the sum by the number of observations. Using the example from the previous section about the commuting times to Baruch over the past 30 days, let's calculate the mean:

Day	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Time	55	57	40	41	44	45	29	45	44	46	48	35	15	39	41
Day	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Time	44	45	46	48	44	47	28	47	42	73	42	45	44	41	40

$$Mean\ (\overline{X}) = (55 + 57 + 40 + 41 + 44 + 45 + \ldots + 73 + 42 + 45 + 44 + 41 + 40) / 30 = 1300 / 30 = 43.33$$

On average, it takes you 43 minutes to get to Baruch. In the section on numeric data representation, we concluded that the most frequent observation was between 35 and 45 minutes, which makes the mean that we just calculated about right and the definition of central tendency true: most values cluster around the mean. However, this is not always the case and sometimes the most frequent observations do not coincide with the calculated mean.

Let's look at another example of commuting times to Baruch. The data you collected over another 30 days will be something like this:

Day	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Time	55	57	40	41	44	45	29	45	44	46	48	35	180	39	41
Day	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Time	44	45	46	48	44	47	28	47	42	73	42	45	44	41	40

By just looking at the data you can say that it still takes you around 44 minutes to get to school most of the time. You will also observe that on day 13 you spent 180 minutes or 3 hours in the train – a very important observation. Calculate the mean:

$$\overline{X} = (55 + 57 + 40 + 41 + 44 + \ldots + 42 + 45 + 44 + 41 + 40) / 30 = 1465 / 30 = 48.83$$

While most observations "cluster" around 44 minutes, an average commute this time is longer. What does it mean? In practical terms, it means that you should not count on the punctuality of public transportation and give yourself only 45 minutes to get to school. Instead, you should budget more time to make sure that you get to your 9:30 class on time. On the other hand, when the Q train is on time, you find yourself roaming empty hallways because you have come to school too early.

In statistical terms, the discrepancy between the mean and the most frequent observation means that the arithmetic average is influenced by extreme values. The last observation that we deemed so important – 180 minutes – is that extreme value. Extreme values can be both on the high end and on the low end. It is high in our case and therefore pushes the arithmetic mean up. Extreme values on the lower end tend to pull the average down.

Arithmetic mean, although a very important measure of central tendency, is often considered inappropriate or insufficient to make judgments or conclusions about the data because of the way it can be affected by single observation.

Median

The median solves the problem of extreme values. The median is a central observation in an ordered array, for which half of the observations are smaller and half of the observations are larger. If you write out the commute times again in the ordered array, you will see that the central value – the median – is 44 minutes.

How did we get 44? To find central value, count the number of observations you have. In our case, it is 30. Logically, the central observation is between the 15th and 16th observations. To find the median, add these observations and divide by two – find the average of the two:

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Time	28	29	35	39	40	40	41	41	41	42	42	44	44	44	44	44	45	45	45	45	46	46	47	47	48	48	55	57	73	180

If you prefer formulas,

Median's position = (Number of observations + 1) / 2 = 31 / 2 = 15.5

The result 15.5 means that median is located right between the 15th and 16th observations. **Please remember that the above formula gives you the position of the median in an ordered array, not the median itself.** In our case, the 15th observation is 44 and the 16th observation is 44. Obviously, the average of the two is 44.

To practice, find the mean and the median cost of your textbooks per semester (check your answers on the next page):

Psychology (used)	$45
Microeconomics (used)	$50
English (3 used books)	$60
Fundamentals of Business (new)	$80
Statistics (new)	$120

$\overline{X}$ = (45 + 50 + 60 + 80 + 120) / 5 = 71

Position of the median = (5 + 1) / 2 = 3

Median (Q_2) = 60 (third observation in an ordered array)

If you pay attention to some statistical findings, you will notice that the median is used much more often than the mean, because the median is not influenced by extreme values. The median household income and the median cost of new homes in the area give you a better idea about that area than the mean income and the mean cost of homes. Take New York State, for example. Because of the few people who make a lot of money and because of the very expensive homes in NYC, the mean will be much higher than the most frequent observation in the state. The median is not influenced by these extreme values, so it is more useful in statistical and economic analysis.