1 Lesson 5 Measures of Central Tendencies

1.1 Introduction

We have outlined the correct process for dealing with data. The steps are

1. Organize the data.
2. Arrange the data.
3. Present the data.

We have spent some time on step 3, but have concentrated on presenting the data graphically. In doing so, we have used found that it is often a good idea to rearrange the data by performing a bit of arithmetic. To illustrate various data sets, we created totals, computed a few percentages, and even created a new measure (deviation from par.) Now we are going to look at other arithmetic processes that can help summarize and present large data sets. This is our first real look at the ideas of Statistics.

One of the first and most common uses of descriptive statistics is the use of averages to describe the central tendencies of a data set. For example, when looking at a data set made up of numbers, it is a natural to ask things like:

- What is the center?
- Where is the data balanced?
- What is a typical number of the data set?

Anyone of these questions is libel to get us thinking about taking the average of the numbers. However, on closer inspection things we realize that are a bit more complicated than that. Consider our class of students; suppose we give them one more surprise test and record the grades:

Name	Test
April	80
Barry	
Cindy	78
David	80
Eileen	30
Frank	80
Gena	75
Harry	55
Ivy	80
Jacob	27
Keri	
Larry	80
Mary	100
Norm	70

The results are rather disappointing. There are some pretty low grades; one high one, and two students did not take the test. Well what was the test average? It is 69.6 if you do not count the two absences, but only 59.6 if you do. Either one of these might be someone's idea of the center of the data. Also no one would miss all the scores of 80 in the list, so that might make it the typical score. Also if we say that of the 12 results in this test, half the students (6) have 80 or above, and half (6) have 80 or below. That make 80 another kind of center of these scores. If we say all 14 count in the results, half the students (7) have 80 or above, and half (6) have 78 or below. That makes something between 78 and 80 a way to measure the center of these scores.

So even in a simple example, there are several ways to measure central tendencies of data. Most common such measures have mathematical names like:

- the arithmetic mean,
- the geometric mean,
- the harmonic mean,
- a truncated mean,
- weighted means,
- the median,
- the mode.

All of these can be reasonably used to denote the central value of a data set made up of numbers measuring specific things. Depending on the situation, one of these might be superior to the others, and one or two might be flat out wrong as a measure of center.

1.2 Mean, Mode and Median.

In this section we will study the three most common measures of central tendency: the mean, the median, and the mode. Actually we have just seen all three. In our new test results, we saw two examples of the mean: 69.6 and 59.6. The mean is the arithmetic average of the grade scores: the sum of the scores divided by the number of scores. The only reason there are two means because we have a (non-mathematical) choice between counting the missing scores as 0 or ignoring them. We noticed that many students got the same score of 80. This is called the mode of the data. It is the number that occurs most often in the data.

We also noted that 80 is the median when we said that half the students have 80 or above, and half have 80 or below. To compute the median of a list of numbers we place the numbers in order, ascending or descending, and identify the middle of the set. If there are an odd number of data points, say $n = 2k+1$, then the median is the number $k+1$ from the top. That means that at least k of the numbers are larger than or equal to the median. But then we also know that at least k of the numbers are smaller than or equal to the median. If there are an even number of data points, say $n = 2k$, then the median is the average of the number k from the top, and the number k from the bottom. This also guarantees that at least k of the numbers are larger than or equal to the median and at least k of the numbers are smaller than or equal to the median. We notice that, like the median, we need to choose how many students to count to compute the median. If we use the number of students who took the test, 12 the median is 80. If we use the number of students in the class, 14, the median is $\frac{80+78}{2} = 79$.

Why might there be three different measures of central tendencies like these? Each has it's advantages; each has its disadvantages. However, we will see later that, in large data sets of the right type, these three measures can be expected to be close to each other. So in cases of large data sets the three different computations yield numbers that are not that different. For now, we will concentrate on the differences between these different computations.

Consider the teacher whose class produced these grades. She wants to represent the results of the test using one number that measures the center of the scores. First, we would expect her to decide not to count those students who did not take the test. That is one problem solved. Next it is sort of clear how to answer, "What was the typical grade?" This has to be 80. After all 5 out of 12 students received this score. Suppose she is asked "What was average?" Most people are familiar with the calculation that gives the average, and that it exactly how we calculate the mean of a data set. The answer is 69.6. Notice, however, that this is not a very accurate reflection of the class' performance. Nine of the twelve student did better than this average. The mean has exaggerated the importance of the low scores, particularly the 30 and the 27. This would be far worse if the entire class were counted, the addition of two zeros pulls the mean down 10 full points.

It is in this respect that the median is better than the mean. The median really does split the class in half. If one or two numbers are much larger or smaller than the others, their actual values have less impact on the median than they do on the mean. In this example adding two scores to the mix moved the median less than one point.

Let us consider a test from a larger class. The test is made up of 20 questions each given a value of 5 points. Suppose that the teacher is asked how the class did on the test. After the test is graded, the teacher could quickly tabulate the scores using a stem and leaf plot using the two digits of the numbers. She could then quickly change this into a class and frequency table:

Score	55	60	65	70	75	80	85	90	95
Frequency	10	1	1	3	3	5	5	10	2

If she does not have a calculator or a computer close by, how can she determine the a central grade on the test? With the frequency table above it should be easy to find the mode. Historically before powerful calculators and computers, data sets very often only appeared as frequency tables like this. Consequently the mode was a more common measure of the center than it is today. It was easy to compute by hand, and that was enough for many applications. The most logical and practical place to start analyzing a frequency table was to look for the datum or data in the set that occurred most frequently. You can easily look at our frequency table and see that one mode occurs at 55% (F) and another occurs at 90% (A). In this data set, there is more than one mode. If you tell the students that the center grade on the test was both an A and an F most students would want more information; although this mode information is telling in itself. The mode is easy to compute, but it may not yield one unique number. Typically, the mode or modes do convey some information about the data, but sometimes they do not.

The next most logical and practical idea is to write the test scores in increasing order and find the middle score to produce the median. Although sorting large sets of data is time consuming, that time was already spent in constructing the stem and leaf plot. With a frequency table, the median is easy to find. There are nine different grades, but of course, this is not the important number for finding the median. There are 40 different scores recorded in this table. This is the number we care about, and $40 = 2 \cdot 20$. We need to find the 20-th score from the bottom. To illustrate exactly what we are doing, we will add a row to the table accumulating the number of tests up to that point

Score	55	60	65	70	75	80	85	90	95
Frequency	10	1	1	3	3	5	5	10	2
Scores $\leq$	10	11	12	15	18	23	*28*	*38*	*40*

The score 20 from the bottom is 80. The 20-th from the top will also be 80. So the median grade on this test is 80.

However, the teacher may still be unsure whether the modes or the median provide the best account of the students' scores. There do not seem to be any high or low scores to pull the average off-center. In this case, the arithmetic mean should provide good information, accounting for every datum in the data set. The mean (or the arithmetic mean) is the quotient of the sum of all of the data and the number of the data in the set. Again it is not important that there are only nine different scores in the table; the point is that the table gives us a total of 40 student scores. Again we make the calculation explicit by adding some arithmetic to the table; this time a row of total points and a column of totals.

Score	55	60	65	70	75	80	85	90	95	–
Frequency	10	1	1	3	3	5	5	10	2	40
Points Earned	550	60	65	210	225	400	425	900	190	3025

Thus 40 students earned a total of 3025 points for a average of 76 (C). The (arithmetic) mean of the data is $\frac{3025}{40} = 76$.

The mean is the most difficult to compute of the three possibility we have considered. However, that is because our data is in a printed frequency table, and our calculations are done by hand. But now a days, data is more likely to be stored in a calculator or computer. In this situation, it can actually take the machine longer to find a median or a mode than it does to compute a mean. Although in all but the very largest data sets, a computer can appear to do anyone of them instantaneously. As we will see in large data sets of the proper type, the three measures of central tendencies can be expected to be fairly close to each other. When this is true, finding the mode would give approximately the same information as finding the mean or the median. Each measure of central tendency has its advantages and disadvantages, but in general we will focus on the arithmetic mean because most of its advantages are mathematical rather than simply practical. These properties can be used to predict unknown values for a large population using methods of sampling.

1.3 Comparing measures of central tendency

We have just seen an example of data where there was more than one mode, and further these modes were sufficiently separated that neither would be a good measure of the center of the data. So certainly one disadvantage of using the mode to determine the center of data is that it occasionally does not work as such. So the question arises, what are the advantages of the mode that keeps around? We have promised that as samples sizes become large, the mean, the median, and the mode will all approach the same value, but that may not be enough to bother computing modes. Still the mode is the fastest and easiest measure of central tendency to calculate by hand for large data sets since it often falls out of the first attempt to organize data into a manageable form, like a stem and leaf plot. On some data sets, the median and the mean are inappropriate measures of center simply some data sets are not made up of numbers.

For example, imagine a teaching evaluation. In several areas such as pedagogy, engagement, content knowledge, etc., teachers are evaluated as "needs improvement", "meets the standards", or "exceeds the standards" perhaps by several different evaluators. So how can we represent an overall summary of such an evaluation. There is no mean to be taken. Oh, one could assign numbers to the three categories; say $3, 2$, and 1 respectively. But one could just as well accentuate the importance of a rating of "needs improvement" by assigning the numbers $4, 2$ and 1. This would allow use to compute a mean, but the meaning of this mean would depend more on the assignment of numbers than the outcome of the evaluation. To compute a mean, you need numbers. The mode, however, can be used even when there are no numbers. If a teacher were to receives 5 rating of "needs improvement", 12 ratings of "meets the standards", and 8 ratings of "exceeds the standards," a reasonable summary of the results is that the teacher "meets the standards" because that is the mode of the data. The largest number of ratings were "meets the standard." Here the mode is definitely an appropriate way to describe performance.

Now these three categories do have a natural order. I think that all would agree that, in mathematical notation:

"needs improvement" < "meets the standards" < "exceeds the standards."

Even thought the data points are still phrases, the implied order in those phrases means that we can probably find a median the set. The teacher who received 5 rating of "needs improvement", 12 ratings of "meets the standards", and 8 ratings of "exceeds the standards" has 25 ratings. With $25 = 2 \circ 12 + 1$ data points, we pick out the 12-th data point from the bottom using the implied ordering. This gives a median of "meets the standards." This agrees with the mode. In this particular example, it would take a rather bizarre set of ratings, for the mode and median to turn out to be different. We could certainly construct such a set of ratings. However, in this particular example, the median and the mode will often coincide and make a strong case that their value is the center value of the ratings.

What about assigning numbers to the ratings and computing a mean? If we say "needs improvement" is a 3, "meets the standards" is a 2, and "exceeds the standards" is a 1, then a teacher who receives 5 rating of "needs improvement", 12 ratings of "meets the standards", and 8 ratings of "exceeds the standards" earns $5 * 3 + 12 * 2 + 8 * 1 = 45$ points and has an average of $45/25 = 1.75$. The lower the average the better the teacher. We might change the order of these points so that higher ratings get the higher numbers. After all the numbers we assigned have no natural association to the ratings, so why not make them more intuitive? We may want to reward ratings of "exceeds the standards" by giving them more value. Thus if we say "needs improvement" is a 1, "meets the standards" is a 2, and "exceeds the standards" is a 4, each "exceeds" will cancel out the value of two "needs." All these possibilities are examples of weighted means. The assignment of number values to data classes is "weighting" and the values are the "weights." In weighted data, the mean of the data is a measure of central tendency, but its numerical value and its meaning as such are directly related to the weighting used.

There is another non-mathematical point to be made about the example above. The measures used on the teacher above, he did not seem to get sufficient credit for having more "exceeds standards" than "needs improvement," and that could definitely be seen as unfair. However, the data in the ratings above must have come from the subjective opinions of the evaluators. It is possible to design a rating system that limits the impact of the natural subjectivity of evaluators. All of these are necessarily part of process that produces the data: multiple evaluators; rating check lists; pre-evaluation coordination. After that, there will still be some subjectivity in the data, but one can hope that the precautions taken do provide limits. Once the data is collected, however, no amount of mathematics can possibly remove the subjectivity from the data. Assigning numbers to the ratings and computing a mean to six decimal places is no more precise a measurement than the mode. Subjective in means subjective out.

But now, back to central tendencies. The median is very useful for data sets with outliers, or extreme values in the data set. The median has the nice property that it measures only the relative size of the data, and the actual size. For example, suppose a charity receives donations of \$2, \$5, \$7, \$10, \$15, \$17, \$25, \$30, \$35, \$100, \$1,000, and \$1,000.

The charity would certainly be misleading people if they were to claim that the mode of \$1,000 had anything at all to say about a typical donation. These two are not only atypical, but they are outliers in size compared to the other donations. These two extreme values not only form a nonrepresentative mode, but the pull the mean of center as well. The average donation may arithmetically be \$187.17, but that is no more a representative of the normal donation than the mode. But we have $12 = 2 \circ 6$ donations were made, the median will be the average of the 6th and 7th donation once the donations are written in increasing order. The median donation of \$21 is closer to a "typical" donation to the charity. The median is a measure of central tendency that is resistant to outliers. In small samples of data such as home prices or incomes, extreme outliers are a common problem. Thus the median typically is used to minimize the effect of multimillion dollar mansions, or incomes of persons like Bill Gates. This property of the median, and the fact that it is generally easy to compute, make it almost just as useful as a measure of central tendency as the mean.

The mean has one advantage over all others measures of central tendency that make it the most common measure anyone uses. The arithmetic mean has mathematical properties that make it much easier to analyze than any other measure. we will see most of these properties in use later, but we can point one out right away. Consider the teacher with the large set of test results:

Score	55	60	65	70	75	80	85	90	95
Frequency	10	1	1	3	3	5	5	10	2

We found that the mean, or average, of these scores is 76, or more precisely 75.625. Now 10 students scored 55 points which was 20.625 points below the mean. Two students scored 95 points, 20.625 points above the mean. If we consider all the scores this way, we construct a table

Score	55	60	65	70	75	80	85	90	95
Away from mean	-20.625	-15.625	-10.625	-5.625	-0.625	5.625	10.625	15.625	20.625
Frequency	10	1	1	3	3	5	5	10	2

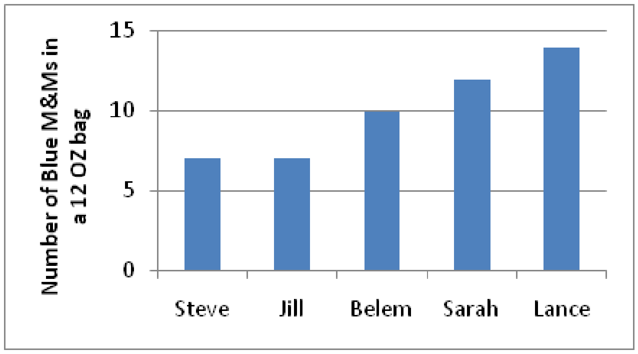
We expand the table to compute the total number of points away from the mean

Score	55	60	65	70	75	80	85	90	95	Total
Away from mean	-20.625	-15.625	-10.625	-5.625	-0.625	5.625	10.625	15.625	20.625	-
Frequency	10	1	1	3	3	5	5	10	2	40
Total away	-206.25	-15.625	-10.625	16.875	1.875	28.125	53.125	156.25	41.25	0

Thus we see that the average distance from the mean is $\frac{0}{40} = 0$.

This is just as easy to see algebraically as numerically, although perhaps a little frightening. Suppose our data is

$$d_1, d_2, d_3, ...d_{n-1}, d_n.$$



The average of this is

$$a = \frac{d_1 + d_2 + d_3 + ...d_{n-1} + d_n}{n}.$$

The distance of each data point from the average is

$$(d_1 - a), (d_2 - a), (d_3 - a), ..., (d_{n-1} - a), (d_n - a).$$

The average of these is

$$\begin{aligned}
& \frac{(d_1 - a) + (d_2 - a) + (d_3 - a) + ... (d_{n-1} - a) + (d_n - a)}{n} \\
= & \frac{(d_1 + d_2 + d_3 + ...d_{n-1} + d_n) - na}{n} \\
= & \frac{(d_1 + d_2 + d_3 + ...d_{n-1} + d_n)}{n} - a \\
= & \frac{(d_1 + d_2 + d_3 + ...d_{n-1} + d_n)}{n} - \frac{(d_1 + d_2 + d_3 + ...d_{n-1} + d_n)}{n} \\
= & 0.
\end{aligned}$$

The mean accounts for every piece of data in the set and finds a mathematical center that can be very pleasing. For example, consider the bar chart below for the number of blue M&Ms that five students collected from a number of bags:

The mathematical mean of the number of blue M&Ms the students found is 10. If we ask the students to redistribute the candy so that each has the arithmetic mean. Now Steve and Jill are both 3 below the mean; while Sarah is 2 above and Lance is 4 above. As we saw algebraically, it is not an accident that the total below is the same as the total above. To redistribute the candy, Belem would stay the same; Sarah would give two, to say Steve; Lance would give one more to Steve, and three to Jill. Because the mean was a whole number, it allowed an equal distribution of the M&Ms to a central value.

One adaptation of the mean that is commonly used is a weighted average. We have already seen one example of this in our teacher rankings. We had teachers evaluated as "needs improvement", "meets the standards", or "exceeds the standards." WE pointed out that data collected this way would have no mean. To compute a mean, you need numbers. One way around this is to assign numbers, or weights, to the three categories; say $3, 2$, and 1 respectively. The teacher who received 5 rating of "needs improvement", 12 ratings of "meets the standards", and 8 ratings of "exceeds the standards" has received a total of 47 weight-points in 25 ratings. This gives a weighted average of $\frac{45}{25} = 1.75$. Remember, the lower a score, the better the rating. This weighted average is 0.12 below 2, and therefore slightly higher than weighting of the center "meets standards" rating.

Now weights like this can be assigned anyway one likes. We could reverse the weights above so that higher numbers are better. If the three categories are weighted as $1, 2$,and 3, then the weighted average is $\frac{53}{25} = 2.12$. This is, not coincidently, exactly 0.12 above 2, and so the weighted average is again slightly higher than weighting of "meets standards." As we change the weights, the weighted average will change. But, as long as our three weights are equally spaced, the results will also give a rating slightly above the center.

Weighting can also be used to add importance to one particular rating. An administrator might choose to give more weight to "needs improvement" by assigning the numbers $0, 2$ and 3. Notice this keeps the same spacing between "meets" and "exceeds,", but stretches the gap between "needs" and "meets." This time the weighted mean is $\frac{48}{25} = 1.96$, which drops the teacher below the 2 level.

Many teachers use weighted averages in a different way on numerical data the data in a set are not inherently worth the same importance. For example, students receive grades of three types: tests, quizzes, and final exam. The grade in each category is computed as a percentage of the category, giving three pieces of data like:

Tests	Quizzes	Final
84%	91%	77%

Rather than simply average these three components to determine a summary grade, the teacher may choose to weight the different components. She might want the quizzes and the final to count equally, but the test average should count 3 times as much. Choosing weights in these ratios will do the trick:

	Tests	Quizzes	Final
Scores	84%	91%	77%
Weights	3	1	1

The weighted average is computed by multiplying the scores times the weights, dividing the sum of the total weights scored, and finally dividing by the total of the weights used.

	Tests	Quizzes	Final	Total
Scores	84%	91%	77%	-
Weights	3	1	1	5
Scored Weights	252	91	77	420

The weighted average is $\frac{420}{5} = 84\%$.

This distribution of weights means that the test score counts as three scores, while the quizzes and the final count as one test each. That means that the test average counts as $\frac{3}{5}$of the total score; the quiz average counts as $\frac{1}{5}$ of the total, and the final as $\frac{1}{5}$.Thus the exact same calculation could appear as

	Tests	Quizzes	Final	Total
Scores	84%	91%	77%	-
Weights	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	1
Scored Weights	50.4	18.2	15.4	84.0

For that matter, $\frac{3}{5}$ is also 60%, and $\frac{1}{5}$ is 20%. So another version of the same calculation is

	Tests	Quizzes	Final	Total
Scores	84%	91%	77%	-
Weights	60%	20%	20%	100%
Scored Weights	5040	1820	1540	8400

Again the weighted average is $\frac{8400}{100} = 84\%$.

Prepared by: Daniel Madden and Alyssa Keri: May 2009