

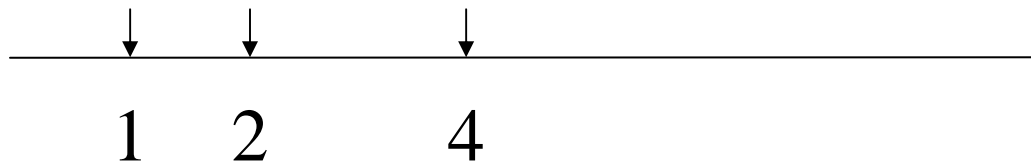
Key Theoretical Concepts

What's In a List of Numbers?

Goal: We study a way of *partitioning* the information in a list of numbers. This partition makes clear a number of facts about the way numbers behave, the way they store information, and the way that information can be modified selectively.

The Number Line Diagram

This is a simple device for visual display of a list of numbers. Using the number line diagram, we can see things that may not otherwise be obvious.



Listwise Operations

We are going to ask a fundamental question: What happens to a list of numbers when we apply the same transformation to every number in the list?

For example, what happens to a list of numbers if we add 2 to every number in the list.

Such an operation, applied to every number in the list, is called a *listwise operation*.

Listwise Operation

Often, we can use a simple equation to indicate a listwise operation.

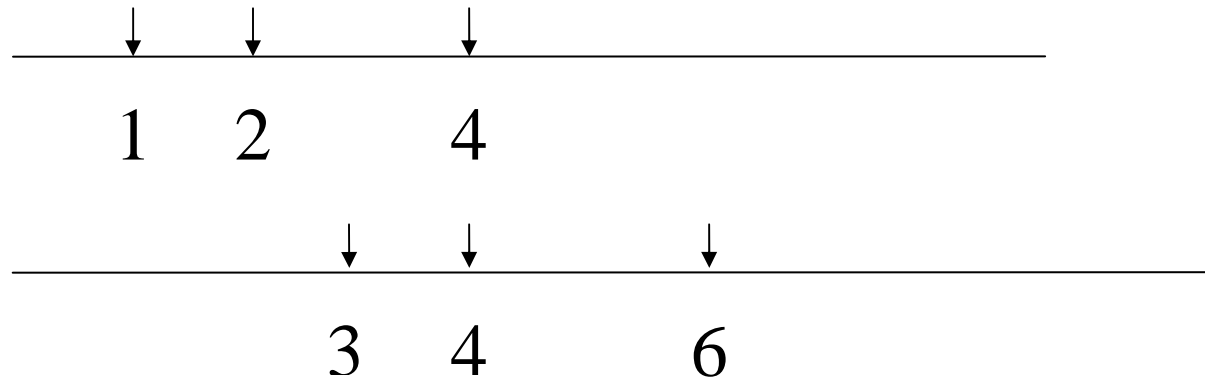
For example,

$$Y = X + 2$$

means “add 2 to every number in the X list to create a new list, called Y .”

Listwise Addition (or Subtraction)

$$Y = X + 2$$



Listwise Addition

Describe, in your own words, how the list of numbers behaved.

Base this description on what you saw in the number line diagram.

What changed about the numbers?

What did not change?

Listwise Multiplication (or Division)

Multiply (divide) all the numbers by a constant. For now, we will restrict ourselves to *positive multipliers*.

$$Y = 2X$$

↓	↓	↓	
1	2	4	
	↓	↓	↓
	2	4	8

Listwise Multiplication

How did the numbers behave? What did you see?

What changed? What remained the same?

Another Listwise Multiplication

$$Y = 2X$$

	↓	↓	↓	
	-1	0	4	
↓		↓		↓
-2		0		8

Summary

Listwise addition or subtraction moves the numbers as a group, as though they were mounted on a rigid stick, and slid to the left or right. Listwise addition does not change any of the distances between numbers.

Listwise multiplication or division by a positive number can move the numbers as a group, but also causes them to “fan in” or “fan out.”

What's In a List of Numbers?

Location

Spread

Shape

NOTE!!!

We shall use simple measures of Location, Spread, and Shape in deriving a number of important principles, but it turns out that these principles hold for *any* reasonable measures of these three quantities.

Location

In what general region is the list located on the number line?

What number is typical of the entire list?

What number is in the center of the list?

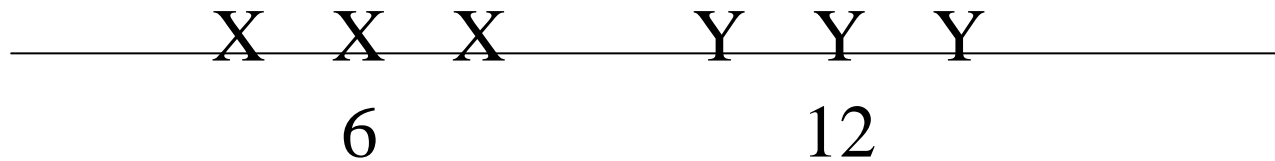
The Median – A Simple Measure of Location

In the discussion that follows, we will use the *median* as our measure of location. There are many measures we could use.

If the number of numbers (N) is odd, the *median* is simply the middle value. If the number of numbers is even, we will define the median as the average of the two middle values, i.e., a point halfway between the two middle values on the number line.

The Median

The X 's have a median of 6. The Y 's have a median of 12. We will use the letter M to stand for the median.

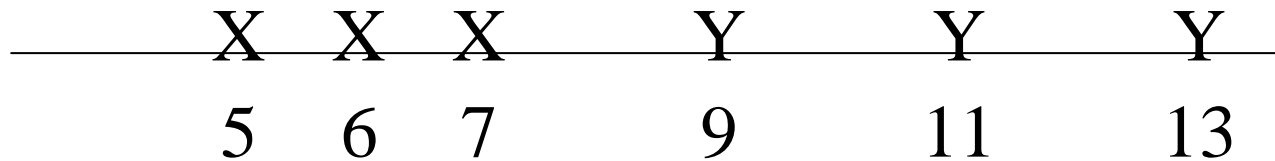


Spread

Measures of spread, or *variability*, assess how far the list is spread out over the number line. In our discussions, we will use a very simple measure of spread, the *range*, which is the difference between the highest and lowest number in the list. We will use the letter S (for *spread*) to stand for the range.

Spread

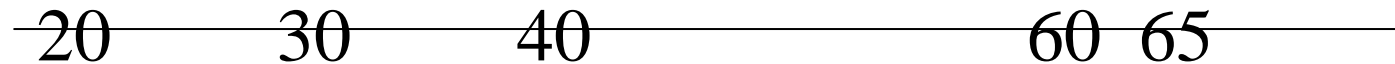
The X 's have a range of 2. The Y 's have a range of 4.



Shape

Shape of a list of numbers is the *pattern of relative interval sizes, moving from left to right*.

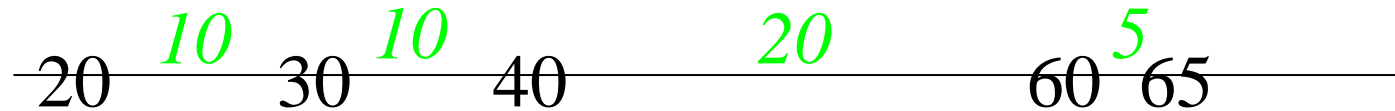
Example. Consider the list 20,30,40,60,65



We can compute the *absolute* distances between the numbers as 10, 10, 20, and 5.

Shape

In the picture below, the *absolute* distances have been inserted above the line between the numbers, in green *italics*.



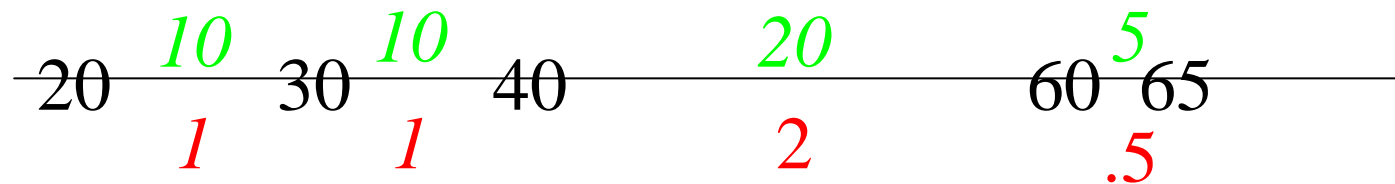
But these are absolute distances. Shape is characterized by *relative distances*.

Shape

To compute the *relative distances* and characterize shape, we divide each *absolute distance* by the first distance, so we re-express all the distances as their ratio relative to the first distances.

Shape

We show these *relative distances* in *red italics* in the drawing below.



Notice that with 5 numbers, there are 4 intervals, and 4 shape parameters (the numbers in red). But since the first non-zero shape parameter is always 1, only 3 of the shape parameters is free to vary.

Shape

To test yourself, answer the following:

What are the shape parameters for

1,2,3,6,10

2,4,6,12,20

13,16,19,28,40

Partitioning the Information in a List

In the list of numbers we just processed, there were 5 numbers. That is, $N = 5$.

We can *partition* the information in the list into 5 new numbers, in such a way that the information does not overlap, but provides us with some key insights. Note that there is 1 measure of location, 1 measure of spread, and there are $N - 2$ shape parameters that are free to vary.

Partitioning the Information in a List

Since $1 + 1 + (N - 2) = N$, we find that the N numbers we started with have been re-expressed in terms of N new numbers.

It is a fact that the original list of N numbers can be reconstructed perfectly from these N new values. (You should try to do this as an exercise).

Partitioning the Information in a List

The information in the original numbers is still there, but it has been re-expressed. Why would we want to re-express the information in a list?

What have we gained?

Effect of Listwise Operations

To understand the value of partitioning information into location, spread, and shape, let's examine the effect of a listwise operation on each of these three measures of numerical information.

Let's start with our simple list of 3 numbers, first calculate the location (M), spread (S), and shape values.

Effect of Listwise Operations

Consider the list of numbers 1,2,4. What is the location?
What is the spread? What is the shape?

Now, here is what I want you to do. Using number line diagrams, I want you to answer some questions about listwise operations.

I want you to answer these questions by drawing a number line diagram, performing the listwise operation, and examining the result *visually*.

Effect of Listwise Operations

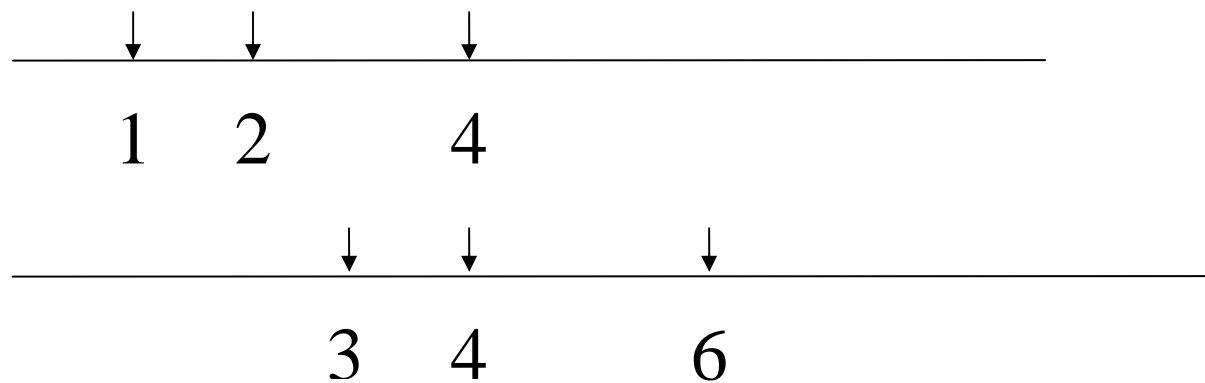
1. What is the effect of listwise addition or subtraction on Location? On Spread? On Shape?

2. What is the effect of listwise multiplication or division by a positive number on Location? On Spread? On Shape?

I'll do the first one for you.

Effect of Listwise Addition

Listwise addition. $Y = X + 2$. Notice that, in the original set of numbers, the location is $M = 2$.



Effect of Listwise Addition

When we perform a listwise addition, all the numbers slide to the right 2 units. So the location has moved 2 units to the right as well. So the new location is $M = 4$.

Unless our eyes are fooling us, we have learned a general rule: *Adding a constant to every number on a list adds that constant to the location. Or, addition and subtraction directly affect location.*

Effect of Listwise Operations

Now let's continue

1. What is the effect of listwise addition or subtraction on Spread? On Shape?
2. What is the effect of listwise multiplication or division by a positive number on Location? On Spread? On Shape?

The Vulnerability Box

<i>Operation</i>	<i>Effect on</i>		
	<i>Location</i>	<i>Spread</i>	<i>Shape</i>
$+$	$+$		
$-$	$-$		
\times	\times	\times	
\div	\div	\div	

Exploiting the Vulnerability Box: Some Examples

1. Tracking changes in a list of numbers.
2. Rescaling numbers. (Changing the location and/or spread without affecting shape).
3. Deriving statistical theory.

Tracking Changes in a List

Any listwise operation that can be expressed as a sequence of listwise additions, subtractions, multiplications, or divisions (by a positive number) can be tracked with the vulnerability box.

Tracking Changes – Effect on Location

Example: You have a list of X 's with a location of 75 and a spread of 20. What will the location and spread become if you convert them to Y 's with this formula?

$$Y = 4 \left(\frac{2X + 30}{20} \right) + 2$$

Location starts at 75. **All listwise operations affect location:**
Multiply by 2 ($75 \times 2 = 150$), Add 30 (180), Divide by 20 (9),
Multiply by 4 (36), Add 2 (38).

Effect on Spread

$$Y = 4 \left(\frac{2X + 30}{20} \right) + 2$$

Spread starts at 20. **Only multiplication or division listwise operations affect spread:** Multiply by 2 (40), Add 30 (**no change** 40), Divide by 20 (2), Multiply by 4 (8), Add 2 (**no change** 8).

Effect on Shape

None!

Interval Level of Measurement

Numbers are used to represent quantities. Often, however, part of the information in numbers is essentially arbitrary. Many kinds of numerical information achieve only what is called an *interval level of measurement*. Roughly speaking, numbers achieve this level *when the shape is meaningful*, i.e., when the ordering and spacing of the numbers properly reflect the order and spacing of the quantity being measured.

Interval Level of Measurement

Consider course grades, for example. Suppose 3 students get grades of 70, 80, and 85. This probably does not mean that the student who got 80 knows exactly 80% of what was taught in the course. Indeed, it might well be that someone who knows 80% of what was taught in the course will get a grade around 90. However, if the grades have achieved an interval level of measurement, the ordering is correct, and the spacing is correct. That is, the difference in performance between a grade of 80 and a grade of 70 is twice the difference in performance between a grade of 85 and a grade of 80.

Interval Level of Measurement

When numbers are at an interval level, we can view the *metric*, i.e., the location and spread, as essentially arbitrary. It is the *shape* that conveys the meaningful information. We want to be careful not to destroy that information.

From the preceding discussion about the “vulnerability box,” we have learned something important. We can alter the metric without affecting the shape, as long as we use listwise addition, subtraction, multiplication (by a positive number, or division (by a positive number).

The Linear Transformation

Any transformation of the form

$$Y = aX + b$$

is a *linear transformation*.

Notice that **any listwise addition, subtraction, multiplication or division can be expressed in the above format. (Why? C.P.)** So the rules of the “vulnerability box” are also “the laws of linear transformation.”

Scaling Course Grades

Suppose you have a set of grades that are at an interval level of measurement, which means that the shape is OK.

However, the class average is 50 and the standard deviation is 20. At Vanderbilt, a more reasonable “metric” for the grades would be to have a class average (mean) of 70 and a standard deviation of 12. How can you rescale the grades without changing the shape?

Scaling Course Grades

The Vulnerability Box tells you that you can alter the standard deviation with multiplication or division without changing the shape. If the standard deviation is currently 20, but we want it to be 12, what do we have to multiply by?

What do we have to multiply a 20 by to turn it into a 12?
Since $12 = 20 (12/20)$ the answer is $12/20$!
Multiply by the ratio of what you want to what you have!

Scaling Course Grades

We can turn this into a formula.

We are going to transform the original scores into new scores with the formula $Y = aX + b$, and now we see that

$$a = \frac{S_y}{S_x}$$

Scaling Course Grades

What happens to our original X 's after we multiply them by .60?

The Vulnerability Box tells us. The standard deviation changes from 20 to 12. But the mean is also multiplied by .6, so it changes to $.6(50)=30$. But we want it to be 70. So what can we do?

Add the difference between what we now have, and what we want!

Scaling Course Grades

We now have $.6(50) = 30$, but we want 70. So we add 40. We can also turn this insight into a formula. Remember, $a = S_y / S_x$, and the mean that was \bar{X} has changed to $a\bar{X}$. We want it to be \bar{Y} , so we have to add

$$b = \bar{Y} - a\bar{X}$$

in this case

$$b = 70 - (12/20)(50) = 70 - 30 = 40$$

You Try It

A class takes a really easy test, and the class mean is 88 and the standard deviation is only 8. The instructor realizes that she has to rescale the course grades so that the class average is 75 and the standard deviation is 12. What formula of the form

$$Y = aX + b$$

should she use?

Putting the Formula Together

We can make one very impressive-looking formula by plugging in our expressions for a and b . We get

$$Y = \left(S_y / S_x \right) X + \bar{Y} - \left(S_y / S_x \right) \bar{X}$$

R Demonstration

Now we'll use some statistical software to rescale some course grades. Suppose the grades are 68,72,80,80,75, and we wish to rescale them to a mean of 80 and a standard deviation of 10.

R Demonstration

Wasn't that fun?

Let's try some more complicated numbers, and see how we can avoid using our calculator.

Z-Scores

Suppose we had taken our numbers and simply subtracted the mean and then divided by the standard deviation, like this:

$$Z = \frac{X - \bar{X}}{S_x}$$

What would happen to the numbers? Use SPSS and find out!

Z-Scores

We could have predicted this, by using the Vulnerability Box to “track” the progression of changes through the expression.

Expression	Mean	Standard Dev.
X	\bar{X}	S_x
$X - \bar{X}$	0	S_x
$(X - \bar{X}) / S_x$	0	1

The Largest Possible Z-Score in a Sample of Size N

Since the “raw scores” can vary from $-\infty$ to $+\infty$, it might seem at first glance that the Z-scores in a sample of size N are also free to vary over the whole number line. But they aren't!

What *is* the largest possible Z-score if $N= 5$? Try to produce an accurate estimate of this value, in the next 5 minutes.

(Hint: Use R to investigate.)

The Largest Possible Z-Score in a Sample of Size N

Actually, the largest possible Z-score in a sample of size N is given by the function

$$\text{Max}Z(N) = \frac{N-1}{\sqrt{N}}$$

The Largest Possible Z-Score in a Sample of Size N

Here are some values of the function

<i>N</i>	<i>MaxZ(N)</i>
2	.7071
3	1.1547
5	1.78885
7	2.26779
9	2.66667
10	2.84605
20	4.24853
30	5.29465

The Largest Possible Z-Score in a Sample of Size N

This means that if you are in a really small class, you cannot achieve a really high Z-score, no matter how well you do!

The Largest Possible Z-Score in a Sample of Size N

Note that, with the aid of statistical software, you can get a pretty good answer to some questions even if you cannot prove algebraically the answer is correct!

Uses for Z-Scores

If two groups differ only on *metric* (i.e., mean and standard deviation) and you want to equate them, you can use Z-scores. Try this on the scores below.

<i>Group A</i>	<i>Group B</i>
56	70
60	80
64	90

Uses for Z-Scores. What have we learned?

1. Z-scores have the same shape as the raw scores, but have a mean of 0 and a standard deviation of 1.
2. Z-scores can be transformed easily into any other *metric* (mean, s.d.)
3. Z-scores can be used to equate groups that have the same shape but different metrics.
4. Z-scores are invariant under linear transformations of the raw scores.