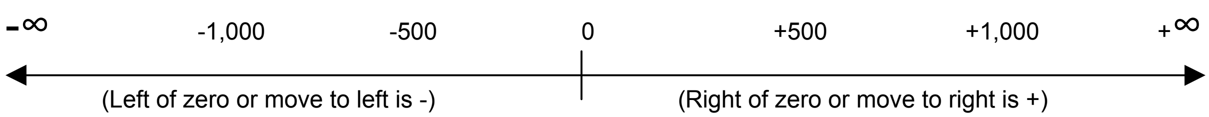
Review of Basic Mathematical Tools used in Statistics

Numbers

Numbers constitute a one-dimensional way of counting things. We can count forward (positively or to the right) or backward (negatively or to the left). The numbers can be represented by a number line that extends from –infinity to +infinity centered on zero (0), which represents the absence of any quantity. All numbers are included on the line, including WHOLE numbers and FRACTIONAL or DECIMAL numbers. 1, 3, 503, etc. are whole numbers, while 1.012357, 3-1/3, and 502.9999999 are fractional and decimal in form. ALL such numbers are included in the number line. Since ANY number has a designated place on the number line, the line is said to be **CONTINUOUS**. Normally, an unspecified number on the number line is known as “x”. Since “x” is unspecified and can represent one or any of the numbers on the number line, we call “x” a variable. The number line is shown below:



The relationships between numbers are laid out in a set of theorems and axioms which, together with the number line, are known as **the real number system**. This system tells us how to deal with the numbers and express how they are related. For example,

3 + 4 = 7 tells us how to add - move from zero to the right 3 units, and then move to the right an additional 4 units.

3 - 4 = tells us how to subtract – move from zero to the right 3 units, and then move back to the left 4 units.

3 x 4 = 12 tells us how to multiply – move from zero to the right 3 units, then move another 3 units, etc., until we have moved 3 units to the right 4 times. We will be resting on the number 12 when the process is complete.

12 / 4 = 3 tells us how to divide – starting at 12, move to the left in groups of 4 units. The number of times required is 3. Hence, division is just the opposite of multiplication. For, say 12 / 5, we must move to the left in increments of 5 twice, which gets us to 2 and then another 2/5ths of a unit. So 12 / 5 is 2-2/5 or 2.4 expressed as a decimal.

From these basic rules are developed further concepts, such as

3 + 4 = 4 + 3, or in a general case a + b = b + a, a and b being real numbers.

3 x 4 + 3 x 5 = 3(4 + 5), or a x b + a x c = a(b + c)
Notice that this requires us to start and 0 and move to the right in groups of 3 for 4 times, and then starting there we must move to the right in groups of 3 for 5 times. When we are done, we find ourselves at 27 on the number line. To get to 27 on the

number line, we could also have started at 0 and moved to the right in groups of 3 for 9 (4+5) times. Hence, the generalized equation above.

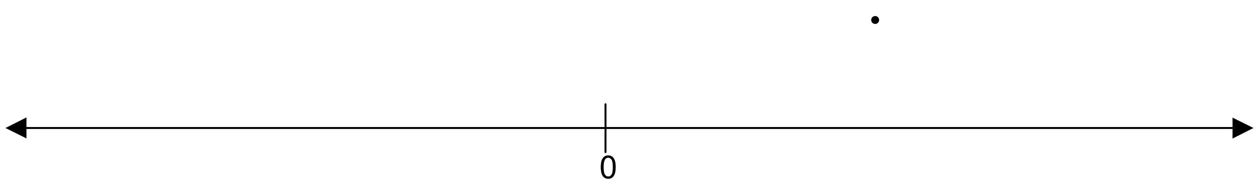
Many of these rules have been developed and are familiarly known as **arithmetic**.

Example: Let μ be a point on the number line and x be another point on the number line. Then x is $x - \mu$ units from μ. If $\mu = 10$ and $x = 15$, then x is 5 units from μ ($15 - 10$).

Mathematics

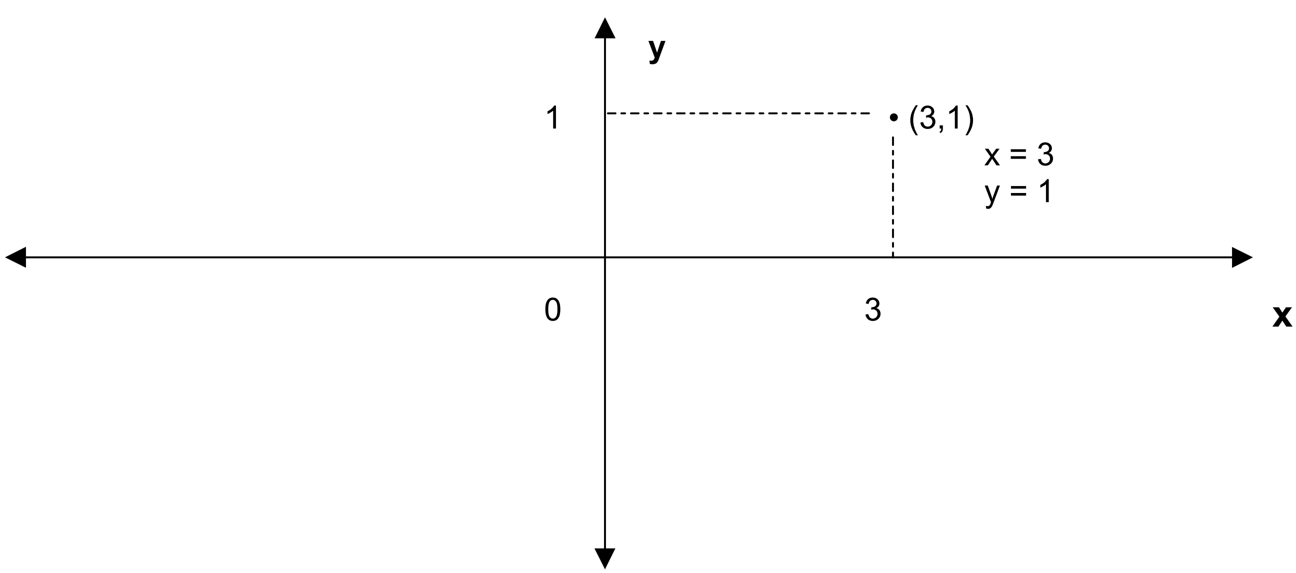
Arithmetic is great for such one-dimensional tasks as counting money or other objects. But nearly all of the real world is not one-dimensional, but has two or three (and even more) dimensions. You can see that it is impossible to apply the number line in its current form to a two- or three-dimensional problem. Let's consider a two dimensional problem:

Consider the number line and the point below:



How do we describe this dot on our number line, given that it is not on the number line?

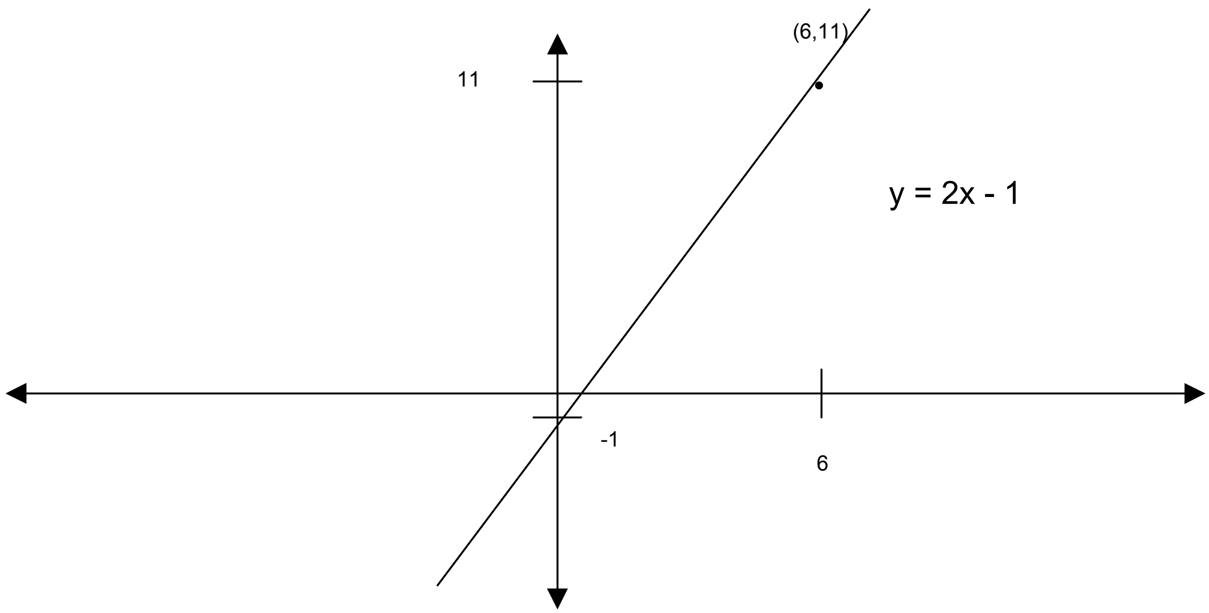
The solution is to intersect our number line with another perpendicular number line that touches the first line at zero. To distinguish it from the original number line, we'll denote values on this second line not by "x", which we reserve for the original line, but by "y". We can now describe the point in question if we can tell its value along both the x and y number lines!



If there is a “line” (straight or curved), it represents a set of points that touch each other. In order to describe the line, we must describe the relationship between “x” and “y” for each point on the line (an infinite number of points). This relationship is described by a formula. The need to describe two or greater dimensional objects with a one-dimensional number line has thus given birth to the field of mathematics – the study of the location of points and their relationships to the various number lines.

For a straight line, the relationship between x and y is expressed as $y = mx + b$. The equation, or relationship between x and y, says that the value of any point on the line along the y-axis is found by selecting a point on the x-axis, multiplying it by a constant amount known as the slope of the line (m), and then adding another constant amount (b), which is the value of y where the line crosses the y axis (y-intercept).

Hence, for a line with slope of $m = 2$ which crosses the y-axis at $b = -1$, the value of y when $x = 6$ is $y = mx + b = 2 \times 6 + (-1) = 12 – 1 = 11$. The location of this point is, thus, (6,11).



The concept is expanded to three dimensions by using a third number line (axis) called z, so that for a point in any space, its relationship to a starting point 0 and any two other variables x and y might be expressed as

$$z = ax + by$$

or some other formula which describes the physical situation. Hence, z is found by knowing x and y and z’s relationship to x and y expressed by the formula $z = ax + by$.

Of course, the relationships between numbers don’t have to be represented by straight lines. For example, the formula

$$y = e^{-(x-\mu)2/2\sigma 2} / \sigma\sqrt{2\pi}$$

shows the relationship between x and y for the familiar bell-shaped curve, known is statistics as a "normal distribution".

The rules for manipulating equations which show numerical relationships between x, y, and z are know as **algebra**. The real number system and algebra together form the foundation of all mathematics. In mathematics, the relationships between variables are called **<u>functions</u>.** These functions are expressed or written using **<u>equations</u>**.

Examples:

If the average age of residents in nursing homes is μ = 77.5 years, and a random sample of 50 nursing home residents has an average of x = 80.25 years, by how many years does the sample average differ from the true population average?

The "how many" or "how far" question is answered by taking the difference between the two numbers:

$\bar{x} - \mu = 80.25 - 77.5 = 2.75$ **years**

Given that μ and σ are constants and x and z are variables, solve the following equation for z: $x = z\sigma + \mu$

To solve for z, you must get z all by itself on one side of the equals sign.

Step 1: Subtract μ from both sides: $x-\mu = z\sigma$

Step 2: Divide both sides by σ: $(x-\mu)/\sigma = z$

Step 3: Rewrite with the solved value on the left: $z = (x-\mu)/\sigma$

In the equation for z above, assume $\mu = 100$ and $\sigma = 10$. An x is randomly selected and has a value of 20. How far is x from μ?

$x - \mu = 20 - 100 = -80$

How far is x from μ per unit of σ?

To find how much a value is per unit of another value, you must divide the two numbers:

$(x - \mu)/\sigma = -80/10 = -8$.

This says that there is -8 units of the value $(x-\mu)$ for each unit of σ.

Given $z = E / (\sigma/\sqrt{n})$, where σ and n are constants, solve for E.

Remember, you must get E all alone on one side of the equation!

Step 1: Multiply each side of the equation by $(\sigma/\sqrt{n})$: $z(\sigma/\sqrt{n}) = E$
Step 2: Rewrite the equation to put the solved value on the left and simplify:

$$E = (z\sigma)/\sqrt{n}$$

Solve the equation above for n.

Step 1: Multiply each side of the equation by $\sqrt{n}$: $E\sqrt{n} = z\sigma$

Step 2: Divide each side of the equation by E: $\sqrt{n} = (z\sigma)/E$

Step 3: Square each side of the equation to get n by itself: $n = [(z\sigma)/E]^2$

Work Practice Set 1

Other Important Mathematical Concepts

Exponents

Exponential notation is common in mathematics and statistics. It is a short-hand way of saying that a number must be multiplied by itself a certain number of times. The number of times is the value of the exponent. For example, if the number 3 is to be multiplied by itself 4 times, the exponential notation would be 3^4. This is stated as 3 to the 4th power. It is calculated as follows:

$3^4 = 3 \times 3 \times 3 \times 3 = 81$.

In general form, a real number x multiplied by itself n times would be written x^n, where x is any real number and n is an integer.

Since x^0 cannot be evaluated, it is defined as 1. Hence

$x^0 = 1$ in all cases. That is, any number to the 0 power is 1.

Evaluate 5^6.

$5^6 = 5 \times 5 \times 5 \times 5 \times 5 \times 5 = 15{,}625$

If n is a negative integer, then $3^{-4} = 1 / 3^4 = 1 / (3 \times 3 \times 3 \times 3) = 1/81 = .0123$.

In general terms,

$x^{-n} = 1 / x^n$

Further, since a negative number times another negative number = a positive number, we know that $(-x)^2 = (-x)(-x) = x^2$. Hence $(-6)^2 = 6^2 = 36$.

Square roots

The square root of a number is the opposite of the square. The square root of a number is a number which, when multiplied by itself, would give us the number for which we desire the square root. Hence, $\sqrt{36} = 6$ since 6 x 6 = 36. In general form $\sqrt{a^2} = a$. Every positive number that can be put under the square root sign is the square of some other number. Note that it is not possible to take the square root of a negative number, since there is no number which, when multiplied by itself, gives a negative number. The square root of a negative number is said to be imaginary and is a subject covered in math texts.

An approximation technique for finding square roots is as follows. Let x_n be a guess of the square root of a given number A. Then, the next best guess for $\sqrt{A}$ is given by the formula

$x_n = 0.5 (x_{n-1} + A/x_{n-1})$

For example, to find $\sqrt{11}$, first make a guess. We know that 3 x 3 = 9, and 4 x 4 = 16, so, $\sqrt{11}$ must be between 3 and 4. Let's select 4 as a starting point. Hence, $x_1 = 4$Then our second guess, x_2, will be found from the formula above.

$X_2 = 0.5 (4 + 11/4) = 0.5 (4 + 2.75) = 0.5 (6.75) = 3.375$

Our third guess x_3 then becomes

$x_3 = 0.5 (3.375 + 11/3.375) = 0.5 (3.375 + 3.2592) = 0.5 (6.6342) = 3.3171$

Likewise,

$x_4 = 0.5 (3.3171 + 11/3.3171) = 3.3166$

$x_5 = 0.5 (3.3166 + 11/3.3166) = 3.3166$

Notice that guess 4 and guess 5 give the same result. Hence, we can say that $\sqrt{11} = 3.3166$. What does your calculator say? Mathematicians have shown that and excellent approximation of the square root of any number is achieved no later than when x_5 is calculated.

The method of determining the square root is called the iterative method.

Work Practice 2

Factorials

Factorials are very important in the field of statistics. The factorial of an integer n is defined by the following formula:

n! = 1 x 2 x 3 x …… x n

Examples of factorial calculation:

3! = 1 x 2 x 3 = 6
6! = 1 x 2 x 3 x 4 x 5 x 6 = 720

Because any number multiplied by 1 is still the same number, the 1 is sometimes omitted in writing out solutions. Also, it is sometimes helpful to write the multiplication chain backward, as in 6! = 6 x 5 x 4 x 3 x 2 x 1, which is the convention that will be used in the statistics class.

Factorials can be treated just like any other real numbers. For example, the factorial (7-4)! Is the same as 3! = 6. Evaluate the expression in parentheses before applying the factorial.

The same would apply to multiplication and division.

7! X 4 ! = (7 x 6 x 5 x 4 x 3 x 2 x 1) x (4 x 3 x 2 x 1)
= 5,040 x 24
= 120,960
while
7! / 4! = (7x6x5x4x3x2x1) / (4 x 3 x 2 x 1)
= 5,040 / 24
= 210.

This can also be written

7! / 4! = (7 x 6 x 5 x 4!) / 4!

According to the rules for real numbers, the 4! In both the numerator and denominator cancel each other, and we are left with 7 x 6 x 5 = 210. This technique is especially useful in evaluating large factorials, as most calculators cannot evaluate factorials for numbers 70 or greater.

As with exponents, we see that the definition of the factorials does not allow for the evaluation of 0!. Hence, by definition,

0! = 1

Example: Calculate the value of x where x = 120! / [(120 – 118)! 118!]

x = 120!/[(2!)(118!)

Notice that 120! can be written as 120 x 119 x 118!

x = (120)(119)(118!)/[(2!)(118!)]

Now the 118! in both numerator and denominator cancel out (=1) because any number divided by itself is 1.

x = (120)(119)/(2!)

Now 2! = (2)(1) = 2, so the solution is just (120)(119)/2 = 7,140

Ratios

A ratio is defined as the division of two numbers. It is usually described as so many units of something per 1 unit of something else.

For example, if a company has $100,000 of sales and 55 employees, the ratio of sales to employees is $100,000/55 = $1,818.18 per employee.

If an egg farm produces an average of 3,000,000 eggs a year and has on average 250 chickens, the ratio of eggs to chickens is 3,000,000/250 = 12,000 eggs per chicken.

Remember that a percent is the number of interest divided by the total number. For example, in a class of 50 students, 6 receive A's. The percentage of students receiving A's is 6/50 = 0.12 or 12%. Note that a percentage is just the division of two numbers and is thus a ratio. If we randomly selected a student from the class, what is the probably that the selected student received an A?

Probability of A = P(A) = 0.12 or 12 out of 100

The Σ function

Whenever the symbol Σ appear before an expression, it means that all the solutions of the expression are to be calculated and those outcomes totaled. For example:

Σx for x = 1, 2, 5, 10 is
$\Sigma x = 1+2+5+10$
$\Sigma x = 18$.

Frequently, the expression following Σ is a formula or function. In such a case, it is necessary to evaluate the formula for each possible value of x and **<u>then</u>** add up the results.

Example: Let G_i be the grade received on test i, where i = 1 is the first test, i = 2 is the second test, i = 3 is the third test, and i = 4 is the fourth test. Then G_1 is the grade on the first test, $G_1 = 85$, G_2 is the grade on the second test, $G_2 = 78$, G_3 is the grade on the third test, $G_3 = 88$, and G_4 is the grade on the fourth test, $G_4 = 70$. Then

$\Sigma G = G_1 + G_2 + G_3 + G_4$
$\Sigma G = 85 + 78 + 88 + 70$
$\Sigma G = 321$

What is the average grade for the four tests?

Average grade = $\bar{G}$ = 321/4 = 80.25

How would you write the calculation of the average grade as a formula?

$\bar{G} = \Sigma G/n$ where n = the number of grades (G's) that are added up.

Work Practice Set 3

Calculation Tables

When the Σ function is encountered, it frequently causes us to make multiple calculations and then do further calculations on the results. This can make it difficult to keep track of and properly complete all the required calculations.

To assist in this process, it is best to put the calculation in the form of a table. For example, suppose it is necessary to find Σxy^2 and we know that y = 2 when x = 3, y = 1 when x = 4, y = 6 when x = 5, and y = 9 when x = 6. How do we calculate this sum. A table is ideal.

x	y	y2	xy2
3	2	4	12
4	1	1	4
5	6	36	180
6	9	81	486
			682

Thus, $\Sigma xy^2 = 682$.

Anytime such a chart can be used, it will reduce the possibility of error because it orders the arithmetic and assures all calculations are completed in the correct order.

Work Practice Set 4

Look-up Tables

Many complex mathematical calculations have already been performed for us and the results are summarized in a table. In such a case, for any value of a specific variable, the corresponding value of the other variable can be found by looking it up in the table. The table has already solved the equation for the desired value of the variable.

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023

The table above is called the Standard Normal Distribution table. It has already solved the normal distribution formula for the special case where $\mu = 1$ and $\sigma = 0$.

When z = .53, what result do you get from the above table?

Look under 0.5 in the left column. The 0.03 is found in the column so labeled. Where the two intersect, the result 0.2019 is found.