**PRAGMATHEMATICS**

***Leveling-up mathematics for empirical (data-based) researchers.***

Some aspiring researchers need to remedy a weak mathematical basis;   
others may feel a need to refresh what they learnt some time ago at secondary school.   
  
The items below are not intended as a course in mathematics, but rather as a repertory of elements   
that students must grasp in order to be able to follow a course of data analysis   
and to understand the typical **univariate** (with one variable), **bivariate** (with two variables) and **multivariate** (with several variables) models and the related statistics, used in research.  
Instead of building up mathematical knowledge systematically and step by step,  
we will often start from concrete issues.   
Often, also, insights will be introduced which will be useful or practical   
in further demonstrations or applications;   
there will therefore be a lot of cross-references between topics in this course   
and repetition of the same or similar concepts, but in different applications.   
  
The focus is pragmatic; hence the title: **‘pragmathematics’.**This ‘refresher’ should be of use (but not sufficient) to prepare   
for general admission tests for graduate studies (e.g. GRE, EXADEN, GMAT).  
 **If you feel familiar with the content of this ‘pragmathematics’,  
then skip this course and move on to other modules of the program.  
But verify first that you are really at ease with all this material!  
Do not avoid the effort now; you may regret that later.**  
  
Where possible, we use Excel to illustrate concepts and carry out computations.   
Today, being able to use Excel is almost a precondition for becoming a researcher.  
Being conversant with Excel will also help you grasp  
abstract mathematical and statistical concepts more directly.

It is important that you not only understand the concepts mentioned below,  
but that you also become used to them, that they become ‘habitual knowledge’.  
We suggest that you study this material first   
with the help of the accompanying video fragments,  
and, next, that you review these notes **three times** on your own.   
By the third iteration, you will start feeling ‘the habit’.   
You will also gain understanding and become habituated   
by doing the exercises in Excel provided along with the video sections.

**An etymological note.**  
Etymology is the study of the origin and meaning of words.  
Two words that we will frequently mention are mathematics and statistics.   
Where do these words come from?   
 *‘Mathematics’* comes from the Greek *máthēma*, which, in ancient Greek,   
means "that which is learnt", "what one gets to know",   
hence also "study" and "science", and in modern Greek, just "lesson".   
To the ancient Greeks, mathematics was what a learned, scientific person should know.   
In that sense, this text is also ‘mathematics’, something you really should know.

Think of mathematics as a language.   
Language is used to describe, to convey meaning, to communicate; to reason;   
language allows some operations, subject to rules (grammar):   
some operations are grammatically permitted, others not;   
linguists study these rules and use them,   
e.g. to computerize and automate language production, translation, etc.  
  
Children are born with a capacity for acquiring a mother tongue,   
even several ones, if they have intense interaction in a language with a parent.  
Other languages can be acquired later, but that requires much effort and pain.  
Mathematics is one of those other languages, one with very precise meaning and rules.  
It may help to think of the learning of mathematics   
as the learning of another language than your mother tongue(s).  
  
‘Statistics’ is related to the word and concept of ‘state’.   
In the 17th century, several states took shape in Europe: France, Spain, England,…   
the rulers of these states were interested to know   
the number of people and of homes in their realm, etc.   
The larger these numbers, the more powerful the state:   
more people meant larger armies, more taxpayers, etc.   
It therefore became necessary to count these numbers, or at least to estimate them.   
Hence the origin of the word ‘statistics’,   
the science and numbers related to the ‘state of the state’.  
Often, these numbers remained uncertain estimates,  
or it was necessary to rely on a sample from the population   
rather than on a fully counted census.  
Hence the association of statistics with the ‘mathematics’ of uncertainty  
and of inferring properties of a whole population  
from observing only a sample of population members.

1. **Introducing you to working with Excel**

If you are conversant with Excel, then skip this section.  
  
If not, the following will be of use, not only for this course,  
but for many things you may need to do later on in many different areas.

Excel is a ‘spreadsheet’ or ‘worksheet’.  
It is meant to help you organize information, data,   
(data is Latin for ‘that which is given, the information you receive)  
and carry out operations on it, especially repetitive ones.

Open the Excel program to find a spreadsheet ready for your work.  
At the bottom, you can open (and name) more than one such sheet.  
  
Familiarize yourself with the rows, columns and cells of a spreadsheet.  
Add (and next delete) a row and a column.  
Adjust the width of a column.  
  
Explore the content of the different toolbars: Home, Insert, Layout, etc.  
Return to the ‘Home’ toolbar.  
Enter text in a cell  
Enter an integer number in a cell.  
Enter a decimal number in a cell   
(find out if it is written with a point or a comma;   
both are possible, but have to instruct Excel to use one or the other).  
Enter a formula (e.g. ‘= 2\*3’) in a cell and verify that the result is correct.

**NOTE: in Excel and in this text, we will use the symbol \* to represent multiplication.  
The symbol x is used to represent the letter ‘x’ and not multiplication.  
Often the multiplication symbol will not be shown,   
e.g. ‘2a’ actually means two times ‘a’, i.e. 2\*a; ax stand for a\*x, i.e. a times x.**  
Enter a formula using the content of another highlighted cell as an argument  
e.g. ‘place in cell B1 two times the content of cell A1’ will be ‘= 2\***A1’**  
Learn to copy a formula down a column.  
  
Learn to add a series of numbers using the ‘autosum’ instruction (upper right of toolbar).

Just so that you know:   
you can perform an amazingly large and diverse set of operations in Excel.

1. **Introductory concepts: numbers, variables, models, parameters.**

Much of what we will discuss deals with numbers, variables, models and parameters.   
What are these key concepts?  
  
**1.1 Numbers**

The most elementary function of mathematics is as a language to deal with quantities:   
the number of instances of individual things   
(e.g. how many **pupils** in a class?),  
or a quantity of things which cannot be counted individually   
(e.g. how much **liquid** in a container?).

Language is an invention of the human mind, and so is mathematics.   
Its words are numbers;   
its grammar are the operations we can perform on numbers (e.g. adding up).  
  
There is nothing sacred about this language;   
like other languages, it has weaknesses and strengths.  
The strength of mathematics is that it is a very a precise language.

But even then, there are several languages in mathematics too,  
with different strengths and weaknesses.  
The Romans used a different numbering systems than us:  
2014 in Roman numbers is MMXIV.  
The Romans did not have a symbol for zero.   
One consequence of such a system is that Romans could not deal with large numbers  
such as 3 565 440 000  
or with small numbers such as 0,12433,  
and that they could not develop ‘counting machines’,  
such as the odometers in our cars (counting the number of kilometers traveled)  
or calculating machines (such as our pocket calculator or Excel)  
or security systems with access numbers, …

Our numbers system originated in ancient India  
and was perfected by Islamic culture;  
it is based on ten symbols: 0,1,2,3,4,5,6,7,8,9

We are all familiar with numbers:   
we use them when we pay in the store, to estimate our time of arrival,  
to read and understand our electricity bill, etc.   
  
Numbers show something of the power of the mind:   
they are the result of our thinking (as are words and grammatical rules).   
Numbers do not really ‘exist’; there is not something   
like a palpable ‘2’, or a ‘34’, or any other number out there in the world,   
like there is, for example, a water molecule or a volcano  
(and even then: there is no such thing in the real world as ‘the volcano’;  
there are only volcanoes, each one different from, though also similar to others).   
  
Numbers are a fiction of our mind; we use them all the time.   
Even those who ‘do not understand mathematics’ use them constantly;   
they are mathematicians without knowing it.   
We cannot function in today’s world without numbers,  
to check the speed of our car, the speed limit on the road,   
the prices of items in the store, the pages of the novel we are reading, etc.   
  
Numbers may be **integers** (or ‘whole numbers’), such as 1 or 2 or 233 or 4671   
or **fractional** numbers, also called **decimal** numbers   
(such as the fraction ¼ or 0,25 or the fraction 3/2 or 1,50).  
When we write, for example, 3/8 that actually means “3 divided by 8”.   
A number divided by some other number, like 3/8 or a/b   
(where a and b can represent any number)  
is called **‘a fraction’**.  
   
**Note: we will write fractional numbers with a comma,  
(as is or was the custom in Europe);   
many people (e.g. in the U.S.A.) write them with a point instead of a comma;   
that is confusing, but we have to live with it!**  
  
Some fractional numbers have no ‘ending’,   
like one third or 1/3 or 0,3333333.  
  
If four people must contribute equally to a gift of 5 $,   
then each must contribute 1,25 $, since 5/4 : 1,25   
(indeed: 4\*1,25 = 5);   
if only three people contribute, then each must contribute 5/3 = 1,666666…,   
this is a decimal number which ‘never ends’.  
  
Since we cannot use ‘endless numbers’ in practice, we will resolve the problem   
by, e.g, two persons contributing 1,67 and the third only 1,66   
(1,67 + 1,67 + 1,66 = 5).

Numbers are positive, zero or negative:  
**positive** (if I carry 1000 $, but owe 400 $ to you, then I really own only 1000 $ - 400 $ = 600 $),   
**negative** (if I carry 1000 $, but owe 2500 $, I am 1500 $ in debt; I ‘own’ 1000 $ -2500 $ = -1500 $),   
or **zero** (if I own 1000 $ and I owe 1000 $, then my worth is 1000 – 1000 = 0 $).

Positive values are larger than zero, mathematically written as ‘**> 0’**;   
negative values are smaller than zero, written as ‘**< 0’**.

If a number, call it ‘x’, lies between 0 and 1, we write that as **0 < x < 1.**

**1.2 A little algebra: “find the unknown”**  
  
This ‘Pragmathematics’ course intends to impart the mathematics   
that you need to carry out scientific research.  
  
Science is often about finding solutions (‘the unknown’) to a problem  
on the basis of evidence (‘data’; literally: ‘that which is given’).

Algebra (derived from the name of an ancient Arab mathematician, Al Gebr)  
deals with the rules to use in order to find the value of an unknown  
on the basis of the information (‘the knowns’) that are given to you (literally: ‘the data’).  
Algebra is more logical thinking than mathematics.  
  
A (very simple) example of a problem:  
I have a number of dollars in my pocket;   
if I give you five of these dollars, I still have four dollars in my pocket  
  
Problem statement: HOW MANY DOLLARS ARE THERE IN MY POCKET?  
That is the unknown.  
Unknowns are usually represented by symbols like x, y or z (the end of the alphabet).  
The 4 and 5 dollars are the information, the ‘data’, often also called ‘parameters’ of the problem.  
  
The solution of this problem, obviously, is that I carry 9 dollars in my pocket!  
This example is very simple   
Yet, as a child in school, you suffered over this problem!  
And you were taught simple rules of logic to solve it, ‘algebraic rules’  
  
This problem can be written as an ‘**equation**’,   
i.e. a statement that two things must be equal:  
 x – 5 = 4 (what I have in my pocket minus 5 dollars is 4 dollars)  
(x is the ‘dollars in my pocket’, the ‘unknown’; -5 and 4 are the ‘knowns’).

**A first rule of algebra is that you may switch terms   
from one side to the other side in an equation if you change its sign.**  
  
You can switch the ‘known’ number 5 to the other side of the equation,  
changing its sign (from -5 to +5)

What I had in my pocket (the unknown, x) is equal to   
the 5 dollars I give to you and the 4 that remain.  
 you then obtain x = 4 + 5, to find the solution x = 9.  
How proud you were as a kid to be able to solve this!  
  
Now, if I give you this problem to solve **x + 1600 = -400**   
we see that the ‘structure’ of the problem is the same,  
but with different parameters (+1600 instead of -4 and -400 instead of +5).

The solution is

………… x= -1600-400= -2000

To state that type of problem even more generally,  
we can substitute the parameter **a** for the first (given) number (-5 or +1600),   
and b for the second one (+4 or -400):  
(note that one uses the first letters of the alphabet to represent parameters).  
and write out this problem more generally as   
 x + a = b  
and the solution as x = -a + b  
In this solution, you can insert any values of a and b that you need   
and obtain the solution of the problem for those specific parameter values.  
You do not have to think over from scratch   
the solution to each specific problem with the same structure.  
  
That is one of the nice things about mathematics:  
once you know the solution to a particular type of problem,  
you can just apply that without having to solve it again.

**In order to be a scientist and to carry out some logical operations,  
you need to be familiar with a few algebraic rules;   
these are rules that you use intuitively every day to solve simple problems,  
but which it is useful to know explicitly to solve more difficult problems.  
  
The rule ‘when changing a term to the other side of the equation, change its sign’  
is therefore important to know and remember.**

Another such rule applies to fractions, i.e. the division of one thing by another  
(e.g. ¾ or x/5).  
Let us consider the following problem:  
I have a number of dollars in my pocket;   
if I give a third of that amount to you, you will receive sixteen dollars;  
how many dollars do I have in my pocket (x, the unknown)?  
You can write this problem as   
 x/3 = 16  
and solve it for x.   
We can compute in our head that the solution is   
x = 3\*16 , which is 48  
Here we apply the following rule:   
**with fractions, when you want to move a term to the other side of the equation,  
you multiply that other side by the inverse of that term.**   
i.e. x/3 = 16 is changed into x = 3\*16  
(3 is the inverse of 1/3).

Likewise, if the problem is 314/x = 2   
then the solution is 314 = 2x  
and hence x = 314/2.  
  
Again, you can write this problem more generally as a/x = b (where a is 314 and b = 2)  
and the solution then is x = a/b

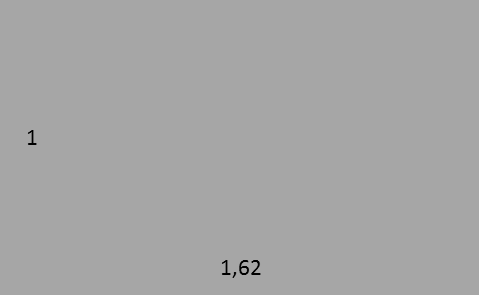
A further useful rule (I don’t really have a name for it),  
is illustrated by the following problem:  
me and my friend set up a company, and we need 10 000$ of capital.  
My friend agrees to invest half of what I invest.  
How much will he have to contribute?  
Let us represent my investment as the unknown ‘x’,  
then his investment will be half of that: x/2.  
  
The problem then is to solve the equation  
x + (1/2)\*x = 10 000$  
This you can write also as   
1\*x + (1/2)\*x = (1 + ½)x = 1,5x = 10 000$.  
This we can solve as x = 10 000/1,5 = 3333,33$  
  
The rule here is:   
if the unknown appears more than once in the equation, each time multiplied by a specific parameter,  
then move the terms with the unknown to one same side of the equation  
(remember to change signs if you change sides)  
and add up the parameters of the unknown (in the example: 1 + ½).  
   
A more general formulation of this problem is  
ax + bx = c (with a= 1 and b = ½).  
The solution is written as (a+b)x = c and hence x = c/(a+b).  
Again, this result holds for any values of a, b and c.

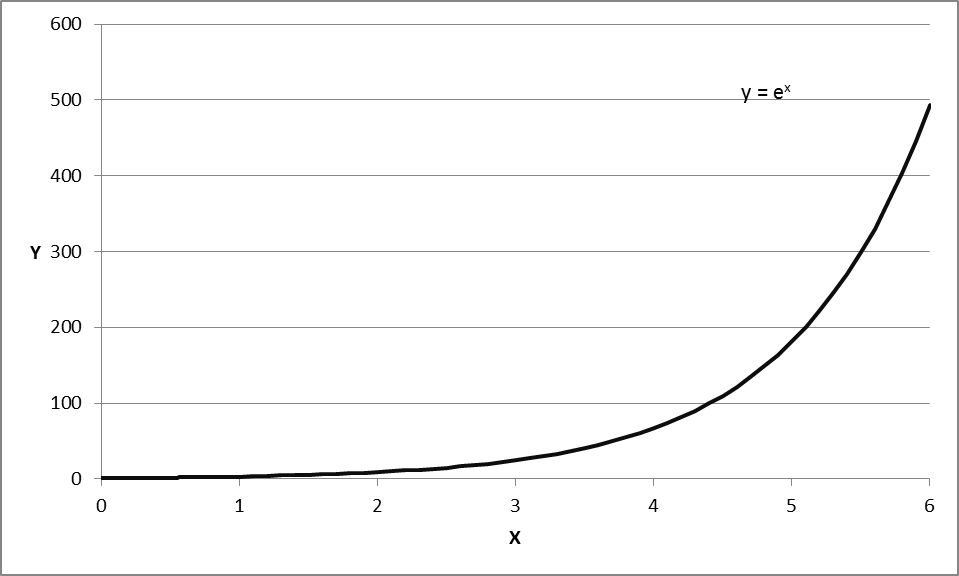
There are many more algebraic rules;   
you have learnt them in high school  
we will remind you of them when we need them.

**1.3 Special numbers**  
  
There are some numbers, often non-integer, ‘endless’ numbers;  
which for some mysterious reason   
play an important role in our life, in the world or even in our universe.   
  
One special non-integer number is ‘***pi***’ (the Greek name for our letter p),   
which equals 3,14159265359…(a number without ending)   
We need ***pi*** to compute the ‘**circumference**’ (the length) or the **surface** (the area) of a circle:   
for a circle with a given radius (‘r’),   
the circumference is computed as 2\*radius\**pi; i.e. 2\*r\*pi (pronounced “two r pi”)*   
the surface is computed as *pi*\*radius\*radius, or ***pi***\*radius² or ***pi\*r*** ² (“pi r square”).   
(Note: we write a number multiplied by itself, e.g. 4\*4, as 4², r\*r as r²).

A consequence is that ***pi*** equals the circumference of a circle   
divided by twice its radius   
(the length of a line from one side of the circle,   
through the middle, to the opposite side).   
  
That is one of the ‘givens’ of our universe, a ‘law of nature’.   
The ancient Sumerians (3000 before Christ) already had knowledge of ***pi***   
and must have used that to build their many circular buildings…

Few will say that circles are not important in our life;   
***pi*** is so omnipresent that we just take it for granted.   
As human beings we are, directly or indirectly,  
using circles (and therefore *pi)* all the time as we go about our daily life:  
the steering wheel of our car,   
the wheels of our bicycle,   
the cooking plates of our kitchen stove,   
the traffic lights,   
the coins in our pocket,   
the turns we make with our car… ,  
all have the shape of a circle, and therefore involve the number *pi*.  
The sun and planets are circle-shaped   
(to be precise, they are spherical: ‘circles in three dimensional space’).  
Many of our sports are based on circles (spheres actually):  
football, volleyball, basketball, tennis, hockey, …

Another special number is the ‘golden number’ or ‘golden ratio’,   
with value 1,61803…   
It is popular in design and architecture,   
where it stands for the proportion between two dimensions of an object   
(e.g. the height and the width of a building)   
which gives humans a pleasurable aesthetic feeling (hence the name ‘golden’).  
  
  
   
Psychological experiments confirm that, given the choice between various rectangles,   
humans tend to find those with proportions respecting the golden ratio most pleasurable.  
Studies confirm that packages of self-service consumer goods  
often respect that ratio, and are preferred for that reason.  
  
Yet another special number is ‘***e***’   
(Euler’s constant or the ‘natural number’), with value 2,71828….   
The mathematician Euler (and others) found it so fundamental to our universe   
(it actually is!) that he called it ‘the natural number’.   
  
We will show below that this number is especially useful   
to describe accelerating (speeding up)   
or decelerating (speeding down) phenomena   
like the one in the picture below.



The natural number ‘e’ will be encountered again several times in ‘pragmathematics’.

**1.3 Systems in numbers: base numbers.**  
  
Since numbers originate in the human mind,  
we can use them to build something systematic.  
This is done by using one number (‘the **base number**’)  
as the central element in a system.

We all know 24, the number of hours in a day.   
For some reason, most people   
have agreed to divide the day into 24 equal parts: hours,   
and the hour into 60 minutes, and minutes again into 60 seconds.   
When we count hours in 24, or two times 12,   
and minutes and seconds in 60, or five times 12,   
we see that there is a **system**:   
with ‘common’ element or ‘base’ number 12.  
  
You may find it strange to count things in a system based on 12   
(a ‘duo-decimal system’).  
Since the French revolution and the era of reason,   
we count most things in a system based on 10 (‘decimal’),   
e.g. length in kilometer, hectometer, decameter, meter, decimeter, centimeter   
weight in ton, kilogram, decigram, milligram…   
volume in liter, hectoliter, deciliter, milliliter, etc.  
  
Still, the system based on 12 persists in some areas   
(eggs are packaged by the dozen or half-dozen in my supermarket;   
the Anglo-Saxon world still measures lengths in inches,   
of which there are 12 to a foot.  
  
We may attribute the persistence of such an ‘old’ system   
to human inertia;   
but there may be more to it than that.  
  
Each base number leads to a system with its own (dis)advantages;   
one advantage of the system based on 12 over that based on 10  
may have to do with how you can divide up a number of things.   
10 things can be neatly divided over 10 people (1 for each), over 5 people (2 for each),   
over 2 people (5 for each) and 1 person (all 10 for that person),  
i.e. in four different ways;   
12 can be neatly divided into 12, 6, 4, 3, 2 or 1,   
i.e. in 6 different ways.  
Hence, it is easier to divide (and share) things   
if you count in twelves rather than in tens...

The real advantage of the decimal system,   
with 10 as base number,  
in addition to the fact that we use ten fingers for counting  
(if we had 8 fingers, the system would probably use 8 as its ‘base’),  
is that measures of different things   
have been integrated in one system using 10 as the base number:  
not only do we measure distances, weights and volumes each   
in a decimal system: meter, kilogram, liter,  
but we have also agreed   
that one kilogram equals the weight of one liter of water  
and that the latter is the volume of a cube with sides of one decimeter.  
These three systems ‘accord’ with one another,   
and that makes life easier.

Sometimes, different systems do not ‘accord’   
with each other or with reality.   
In the Anglo-Saxon world, there is no easy common denominator  
between distances, weights and volumes (feet, pounds, pints),  
and that imposes a lot of unnecessary computations on people.  
  
Another example is the counting of days:   
many societies count days using a system based on the number seven   
(7 days to a week; the seven-day week also dates back to the Sumerians).  
Now, the earth takes 365 days to revolve around the sun.   
Unfortunately, 365 cannot be divided by 7,   
or by any other number except 5 (365 = 73\*5)…  
365 or 73 are not very practical base numbers.   
Some societies counted the days in ‘weeks’ of 12:   
30 times 12 is 360 (almost 365);   
the 5 ‘leftover days’ were devoted to special religious events,   
and were often days during which the population   
lived in awe of this ‘unknown’, this ‘void‘ in time.  
  
To make matters even more difficult, a year is not exactly 365 days,   
but 365 days and 5 hours, 48 minutes, 46 seconds, and a bit more…   
The time it takes the earth to rotate around its axis (the day)  
bears no necessary relationship to the time it takes the earth  
to rotate around the sun (the year).  
Days and years do not ‘accord’ very well.  
The result: we must adjust our calendar based on 365 days per year  
at least every four years (a ‘leap year’ with a February of 29 days).

You see that numbers and number systems are inventions of our mind,   
and that they do not always capture reality perfectly:  
numbers (and mathematics) are only ‘human’.

**1.4 Series of numbers.**  
  
Instead of a single number, we may have a **‘series’ of numbers**.  
Consider, for example, the height (in cm)   
of a class of 8 children shown below  
***(these numbers are only an example; we have invented them ourselves;  
they will be used repeatedly in this course to illustrate concepts):***  
 112  
 115,  
 112  
 106  
 120  
 105  
 111  
 109

This is a ‘series’ of (8) numbers,  
the heights of the children, from the first to the eighth,  
without anything systematic to this sequence of numbers.  
  
Let us represent this series by a symbol, e.g. **W**   
(any other symbol would also be OK!).

When we write such numbers from left to right,   
 112 115 112 106 120 105 111 109  
we represent that mathematically by the symbol **W’**   
(W with an accent ‘, called the **‘transpose’** of W).

Below we show how to perform statistical operations on such series,   
e.g. to compute the average of the heights of these children,   
or the relationship between their height and weight,  
and we extend that to the processing of whole tables of numbers, or ‘data’.  
  
But let us now consider **series of numbers with systematic properties**.  
  
The first we will consider is the **geometric** series,  
   
such as 1 2 4 8 16 32 64 128 256 512 …  
where each number is the previous one, multiplied by 2,   
or, more generally, multiplied by the same number or ‘**ratio**’ ‘r’  
(‘r’ is just a symbol; note: **this is not the ‘r’ of ‘radius’ mentioned above for circles**),   
The series above is called **a geometric series with ratio (‘r’) 2**  
  
Do you know the story of the king who asked the peasant   
how he could reward him for having saved his life?  
The peasant asked for this ‘simple’ reward:

“take a chessboard; put one grain of wheat in the first square,   
then double that in the second square,   
double that again in the third square, and so on,   
until you reach the last square (the 64th) of the chessboard.   
I will take as my reward all the grains on the board.

What he asked as a reward is the sum of the 64 first terms  
of a geometric series with ratio 2   
(the king would not have understood that, he was ‘not good at mathematics’;   
he understood what the peasant asked, but not how much that was…).  
This series grows very rapidly (we will call that **‘exponential growth’** below):  
1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768, and so on…,  
and that is only until the end of the second row of the board.  
The last number (the 64th) is so long that you cannot write it out in normal notation  
(mathematicians will therefore write it as 264, calling that “**two to the 64th**”;  
your calculator will shift to this notation when the numbers become too large for its screen).  
  
When the king reaches the last square of the chessboard,   
he owes the peasant more grain than the whole world can produce…  
  
If this series continues indefinitely, it reaches such large numbers,   
that we might say that it reaches ‘infinity’;   
and then the sum of all the numbers of this series   
must certainly also add up to an infinitely large total.  
  
We see that the sum of the terms of a geometric series with ratio 2 is infinitely large.   
But does the sum of the numbers of any infinitely long geometric series   
always adds up to infinity?  
  
Consider a series with a ratio smaller than 1, e.g. ½ or 0,5:

1 ½ ¼ 1/8 1/16 1/32 1/64 …

i.e. … 1 0,50 0,25 0,125 0,0625 0,03125 0,015625 …  
  
While this series becomes infinitely long, its terms become infinitely small.

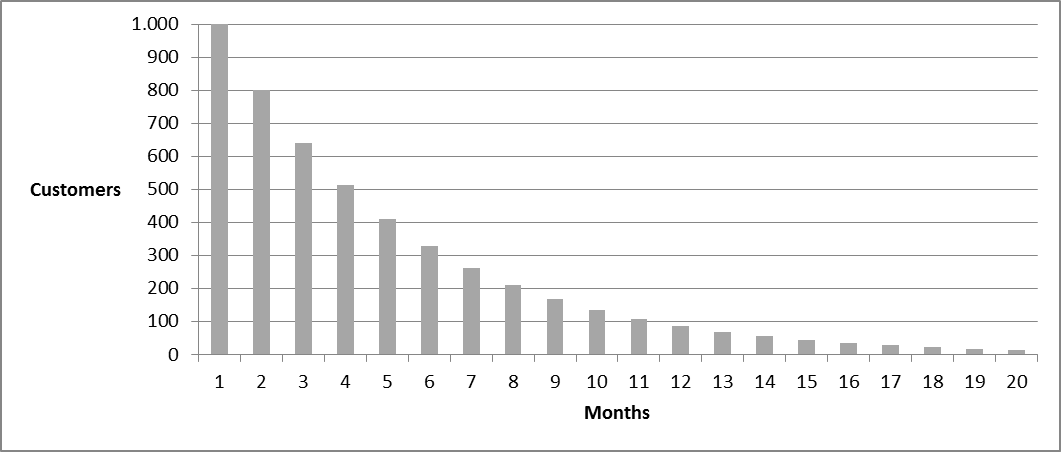
Do the terms of such a series add up to infinity, or to some finite number?  
Now here is a problem that mathematicians relish!   
It allows demonstration purely by reasoning,   
by pure theoretical thinking, by **‘deduction’.**   
  
We will show that the sum of the terms   
of an infinitely long geometric series with ratio smaller than one   
(but larger than zero : 0 < r < 1) is finite.   
  
Let us demonstrate that first for a series with ratio r equal to ½ (r = ½):  
  
i.e. the series 1 1 /2 1 /4 1/8 1/16 1/32 1/64 1/128 1/256 …

Let us use the symbol µ (there is nothing special about µ; it is just a symbol)   
for the sum of the numbers of this infinitely long series  
  
 µ = 1 + 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + 1/64 + …   
  
We can also write µ as 1 + ½(1 + 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + …)  
(to verify this, just multiply the numbers inside the brackets by ½ to obtain the geometric series itself).  
  
Now, an interesting thing appears;  
the sum inside the brackets in the last series is the same as µ;   
therefore, we can write this as   
  
 µ = 1 + ½(µ)   
  
Then, using the rules of algebra (‘change sides, change sign’ and ‘add up parameters of the unknown’):  
 µ – ½µ = 1 or (1-1/2)µ = 1 or (1/2)µ = 1 or µ/2 = 1  
  
and finally, using the rule of algebra for fractions: µ = 2

Conclusion: the terms of a geometric series with ratio ½ (or 0,5) add up to 2!  
  
Now, you could replicate the reasoning above for any value of the ratio ‘r’   
(as long as it lies between zero and one: 0 < r < 1).  
We can therefore generalize the result, the ‘deduction’ above, and write   
   
 µ = 1 + rµ and hence (1-r)µ = 1 and hence µ = 1/(1-r)  
 **Conclusion:  
The sum of the terms of a geometric series with ratio r (0 < r < 1) is 1/(1-r).  
  
This is an important conclusion;   
now that we have given the proof, just memorize the rule, and forget the proof!  
That is an advantage of (prag)mathematics:   
no need to repeat the reasoning once you have given the proof.**

When, e.g. r = 0,8, then µ = 1/(1-0,80) = 1/(0,20) = 5   
  
the geometric series 1 + 0,80 + 0,80\*0,80 + 0,80\* 0,80\*0,80 + … sums to 1/(1-.80) = 1/(.20) = 5.  
  
And when r = 1/3 or 0,3333…, then µ = 1/(1-0,33333…) = 1/(0,66666…..) = 1,5, etc…

If you think that this is ‘pure theory’, without practical relevance,   
then here is a business application that demonstrates one use of this reasoning.  
  
Assume that you are conducting an advertising campaign for a branded product,  
and that your campaign attracts 1000 new buyers in the current month,   
and that 80% of these (i.e. 1000\*80%) buy the brand again next month,   
of whom 80% (i.e. 1000\*80%\*80% or 1000\*64%) buy again the month after next, etc…



Then, how much will the advertising campaign sell in the long run?  
The answer is 1000 (1 +0,80 + 0,80\*0,80 + 0,80\*0,80\*0,80 + ….);   
the expression between brackets is a geometric series with ratio r = 0,80.   
The sum of the numbers in that series, according to the formula µ = 1/(1-r),   
is 1/(1- 0,80) or 1/(0,20) or 5   
In the long run, the ad campaign yields not 1000 1\*1000) additional sales, but 5000 (5\*1000)!   
The campaign may not seem profitable in the short run, but it certainly does in the long run!   
  
Any phenomenon which decreases gradually over time   
according to a ‘geometric decay pattern’, as in the picture above,   
can be represented and studied as a geometric series with ratio smaller than one,   
for example:  
- the extent to which we forget something over time,   
- the depreciation of an asset over time (e.g. the value of your car),   
- the decline of the value of a currency under a constant rate of inflation, …   
  
The geometric series is a useful model to represent and think about such phenomena.  
We will use that knowledge again later in this course.

**1.5 (Series of) Random numbers**

Another useful number concept is that of (a series of) random numbers.

A random number is a number that cannot be predicted.   
If you throw a fair die, the result will be 1, 2, 3, 4, 5 or 6,  
(note: die is the singular of dice; in many games we roll dices)  
but you cannot predict which of these numbers will be thrown:  
a fair die is equally likely to yield each one of these numbers,   
each of these ‘outcomes’ is equally likely.

A (fair) die ‘generates’ integer numbers between 1 and 6 **‘at random’**;  
it is a **‘random number generator’**.   
Other examples of random number generators are   
roulette wheels, innocent children drawing numbers out of a hat,   
machines mixing numbered balls and selecting balls randomly, etc.

If dice are not ‘**fair**’, one says that they are ‘**loaded**’   
(i.e.: they tend to land more on one side than on the other;  
they are not completely random).  
  
If dice (or the draw of the winning number in a lottery) are not fair, there is a problem.  
A few years ago, the ‘random balls machine’ of the Belgian lottery   
got stuck on one specific ball during the draw in full view of the television audience;  
the president of the lottery eliminated that ball from that draw,  
and let the machine draw another ball (another number) instead.  
Not surprisingly, many lottery ticket holders protested that this was not ‘fair’:  
- those with the eliminated number on their ticket wanted that number validated  
- others, who were not winners, wanted the whole draw invalidated…

This shows that random numbers can arouse a lot of passion…

Random numbers, and methods to generate them   
are often encountered in statistics  
(statistics, you could say, is the science or study of unpredictable, random events).

Generators of (series of) random numbers   
e.g. ‘random balls machines’   
may be difficult or costly to realize,   
certainly if you need to generate very large numbers of them.  
  
Throwing dice to generate large quantities of random numbers, for example,   
would not only be very cumbersome,   
but also not satisfactory,   
because not a single die made of matter (e.g. wood) can be perfect:  
dice made of real matter contain some irregularities,  
if only because of the different number of holes or dots on the different sides…

With a computer, i.e. a numbers machine,  
the human mind can build a perfect random number generator.  
One example is the ‘random’ function RAND() available in Excel.  
The instruction ‘= RND()’ will return a random number between 0 and 1  
(how the computer does that is beyond the purpose of this course).  
 *In Excel sheet 1, in cell A1* enter ‘= RND()’ to obtain a random number between 0 and 1.

Let us, as an example, demonstrate how we can randomly   
generate 10 throws of a perfect (‘fair’) die,  
(these are not really the throws of a fair die, but a **simulation** of that process)

Below is a series of 20 random numbers generated by Excel’s RAND() function.

0,8**2**2390 0,2**9**9079 0,7**6**6586 0,5**2**0227 0,8**2**2973 0,4**5**7107 0,7**1**998 0,5**4**5114 0,7**5**6424 0,5**3**8102  
0,6**4**0691 0,4**2**7729 0,2**9**6093 0,2**5**2235 0,7**3**6594 0,1**7**499 0,6**7**2921 0,9**9**4249 0,0**5**0384 0,7**3**2062

How can we use this series to generate (i.e. simulate) 10 throws of a fair die?

We use, for example, the second number, the second ‘**digit**’, after the comma  
of each random number in this series  
(if the number itself is unpredictable, then so will be its second digit);   
we retain that digit if it lies between 1 and 6 (as with the numbers on a die)   
until we have a series of 10 such random digits between 1 and 6.   
This is the same as throwing a fair die 10 times.  
The second digits of the numbers above are   
2, 9, 6, 2, 2, 5, 1, 4, 5, 3, 4, 2, 9, 5, 3, 7, 7, 9, 5, 3.   
Retaining only the first ten numbers between 1 and 6,   
our series of 10 random throws of a fair die would be   
2, 6, 2, 2, 5, 1, 4, 5, 3, 4

This example illustrates the concept of random numbers   
and how these can be generated with the help of a computer.  
This also shows that the computer can be used to ‘simulate’   
a real phenomenon characterized by randomness, by unpredictable forces.   
  
In this case the simulated phenomenon is quite simple: the throws of a fair die,  
an example that you can easily understand.  
  
You should understand that important and complex uncertain phenomena,  
processes with a lot of random, unpredictable influences,   
such as traffic flows, stock market price changes, molecular movements, water flows, etc.   
can be simulated by means of computer programs using random numbers.  
  
Simulating traffic flows, for example, to study how traffic jams develop,  
can be done by imitating the more and less predictable movements of millions of vehicles  
simulated (imitated) by means of random numbers in a computer program.

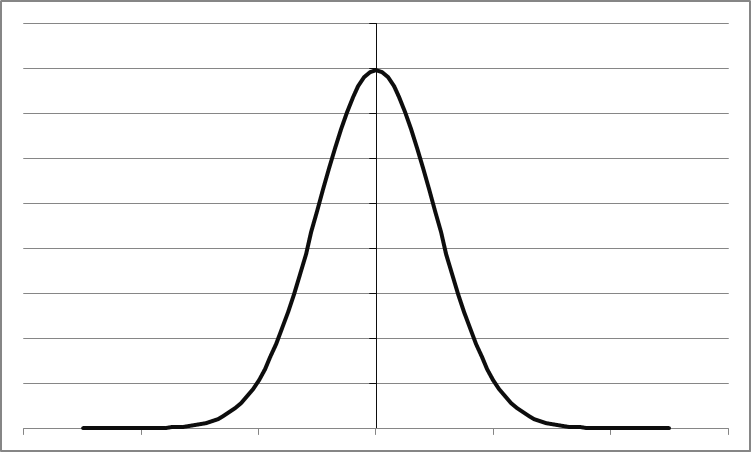
Random numbers and their generation are central to statistics.  
An important task of statistics is to help us decide   
when a phenomenon that we observe   
and that is characterized by a lot of randomness  
is sufficiently strong, sufficiently systematic   
that we must conclude that it cannot be the result   
of only random factors, of only ‘nonsystematic influences’.  
  
There is much discussion, for example, about global warming.  
Weather and temperatures are always quite variable and unpredictable  
and the question is whether the present observations of higher temperatures  
can be attributed to the normal, random fluctuations of the weather,  
or, on the contrary, there is a systematic increase in temperatures going on.  
Statistics will be necessary to answer such questions.

* 1. **Variables**

Variables are phenomena (‘things’) that can take on different values,   
i.e. be represented by different, varying numbers.   
  
For example, my net worth can be 600$, or -1500$ or 0$;   
the (varying) speed of my car could be 50 km/h, 100 km/h, 0 km/h,   
or even -5 km/h (when I back up);   
the pressure of my foot on the gas pedal of my car varies as I drive, etc.  
  
In mathematical language, variables are usually represented by a letter symbol,   
most often by letters (symbols) like x or y or z.  
If x, for example, is used as a symbol for my net worth,   
and my net worth is 600$, we then write x = 600$;   
when my net worth changes and becomes -1500$, we write x = -1500$.   
My net worth x can assume different values; it is a ‘variable’.  
  
When we say that ‘x is a variable’,   
this means that we use the symbol x to represent a phenomenon   
that can take on different values (i.e. that varies, that is variable).

* 1. **Models**

Models are representations of (the properties of)   
a single phenomenon or variable (‘**univariate**’)   
and/or of multiple phenomena and of the relationship(s) between them (‘**multivariate**’).   
  
Models can be expressed in several ways;   
the most important are:   
in words (‘**verbal**’),  
in pictures (‘**graphical**’),   
in mathematical symbols (‘**mathematical**’),   
in **computer code**, e.g. the instructions in an Excel spreadsheet.  
  
For example, the occurrence of the different weights of male adults in a population,   
the ‘**distribution**’ of the variable ‘weight’ over that ‘population’   
can usually be represented well by means of a bell-shaped curve as shown below   
(that is an example of a **univariate graphical model**).



This curve is well known in science and in applied science   
and called a ‘**Normal**’ or ‘**Gaussian**’ curve – after the mathematician Gauss.   
It shows the **distribution** of weights across the population.

The height of the curve indicates how often the corresponding weight occurs:   
weights closer to the average (the middle)   
occur more often than extremely high or extremely low weights.

We understand that such a curve does not show each and every detail   
of the weights of the individuals in the population:   
models are used to simplify,   
to represent only those aspect(s) of a phenomenon   
that we consider important or relevant.   
Here, we use the normal curve to show only  
how the weight of individuals is ‘**distributed**’ over the population.   
  
The **Normal curve** is one of the basic phenomena   
in our world, society and universe, a kind of ‘law of nature’;   
many characteristics are ‘distributed’ over a population   
according to a ‘normal-’ or ‘bell shape’,   
e.g. the weight of adult humans, the wing-span of adult condors,   
the IQ-scores of 10-year old children, the height of waves in the ocean, etc.  
Normal curves will be used often in this and in subsequent courses.

An example of a different type model   
is that of a relationship between two (or more) variables,  
e.g. the statement that

“the gas consumption (in liter/100 km) of my car varies   
according to how deep (in cm) I push on the gas pedal”

can be stated more succinctly as a **bivariate verbal model**  
  
 push on gas pedal (cm) -> gas consumption of car (l/100 km)

A further example is given by the statement

“the gas consumption (GC) of my car (in liter/100 km) equals   
 - 3 liters (the stationary gas consumption)   
 - plus one half liter for every centimeter that I push in the gas pedal (PP)”

This can be ‘translated’ into the bivariate **mathematical model**:  
  
 GC = 3 + 0,5PP

GC and PP are variables.  
3 and 0,5 are (specific) parameter values.  
  
We can then easily compute the value of GC for various values of PP,  
*Excel is an easy tool to carry out such repetitive computations (see sheet 3)*

In the above relationship, the number 0,5 may not apply to every car:   
for some – more economical - cars the number may be lower,   
for other –less economical - cars it may be higher;   
and the stationary consumption of the engine will also differ between cars.  
  
To represent the same phenomenon more generally, therefore,   
we may write the bivariate mathematical relationship as  
  
 GC = a + b\* PP  
  
(where ‘a’ represents the stationary consumption  
and ‘b’ the additional consumption in function of the pressure on the gas pedal)  
we call ‘a’ and ‘b’ ‘**parameters**’; they may be different for different cars.  
GC and PP are variables: they are symbols that represent things   
that can take on different values;  
in this case gas consumption and pedal pressure.  
  
Again, such computations are easily carried out in Excel,   
also for varying parameter values  
*(See Excel sheet 3 for an example of a model in computer code)*.  
  
If we represent the variable GC more generally as y and PP as x,   
the relationship is then written as   
 y = a + b\*x

which is the general way to write a bivariate linear relationship,  
also called **a line** or **a linear relationship**.  
Below we show that if you make a picture of this relationship,  
a graphical model, it will look like a straight line (hence ‘linear’ relationship).

1. **Basic operations on numbers: adding, subtracting, multiplying, dividing.**

You are familiar, of course, with the basic operations of   
- addition (e.g. 2,6 + 3,25 = 5,85)  
This operation is so basic, that Excel provides it as a standard option  
(called ‘autosum’).  
  
To compute the total of a series of numbers,   
e.g. the weights of the 8 pupils given above.  
*Try the ‘Autosum’ function in Excel sheet 1: enter 8 numbers in a column,  
highlight these eight numbers and use the ‘autosum’ function in the upper right of the toolbar.*  
- subtraction (14,34 – 11,12 = 3,22); subtraction can lead to ‘negative’ numbers.   
If, as above, you have in your pocket 1000$, but you have a debt of 2500$,   
then your ‘net worth’ actually is 1000$ - 2500$ = -1500$   
  
- multiplication (2,5\*3 = 7,5)   
  
- division (e.g. eight divided by two is four; and 8/5 = 1.6; and 8/3 = 2,6666666…;   
in the latter case the series of 6’s can be seen to continue indefinitely, without end.   
  
These basic operations are so useful, that we have developed devices   
to perform them efficiently:   
procedures learned at school for complex multiplications or for long divisions,   
pocket calculators, (repetitive) operations by means of Excel.  
  
In Excel, for example, if you have two series of numbers,  
one for the Height and one for the Weight of 8 pupils, H and W.

Height (H) in cm Weight (W) in kg

112 50  
 115 51  
 112 46  
 106 49  
 120 54  
 105 51  
 111 59  
 109 55

and you want to multiply each height by each weight,  
*that is easily done by means of Excel instructions (see Excel sheet 7);   
each of these is called a ‘cross product’ (of height and weight).  
You enter the first cross-product   
as the product of the cells containing the first two numbers  
and copy this formula down to multiply the other seven numbers.*  
A further and useful example of addition and division (see Excel sheet 2)   
is the computation of the **mean** or **average** of a series of numbers.   
*If we want to place a number on the average height of the 8 children,   
we first total (i.e. add up; autosum) their heights (the total is 890);   
then divide (i.e. division) that total by the number of children (8),   
to obtain the ‘****average****’ or ‘****mean****’ of the height of this group: 111,25 cm,*

The **mean** or **average** is a fundamental statistical number   
- to describe a population (e.g. all the people, all the towns, …) or  
- a sample out of a population (i.e. a limited number of people, of towns,…).  
It **describes** what, on average, is the height (or any other characteristic)   
of a population or of a sample out of a population.  
  
**Note: The (importance of the) order of operations on numbers**.  
  
When several operations (additions, multiplications, etc.) must be carried out,   
one must specify the order in which these must occur.   
To compute the average height of a group of children, for example,  
we first total the numbers and only then divide the total by the number of observations.  
The order in which you carry out these operations matters.  
The operation to be performed first, the addition, is placed in brackets  
only next comes the division by the number of observations, 8:  
(112 + 115 + 112 +106 + 120 + 105 +111 + 109)/8  
and not 112 + 115 + 112 +106 + 120 + 105 +111 + 109/8  
(the latter would be 794,625, not a good number for the average height!)  
  
*This is important with Excel,   
especially when computing more complicated expressions;   
the order in which operations are to be carried out must be made clear.*

1. **Exponentiation**

**3.1 Exponentiation of a whole number by a whole number.**

Mathematical writing can sometimes be shortened,  
just like when we use abbreviations in written text.  
A special case of this is when you multiply a number by itself several times,   
e.g. 2\*2 which we write as 2² (or 2^2 or POWER (2;2) in Excel)   
e.g.4\*4\*4 (i.e. 56) which we write as 4³ (or 4^3 or POWER (4;3) in Excel)  
We call this ‘***exponentiation***’: we write the number of ‘self-multiplications’   
as a small number above and behind the self-multiplied number   
When we write 2³ (for 2\*2\*2), 2 is called the ‘base number’ and 3 the ‘exponent’.   
For 2³ we say “**two raised to the power 3**” or ‘**two to the third power**’.  
100 (i.e. 10\*10) is 10² (10^2 or POWER(10;2); 1000 is 10³, etc.   
  
The second power of a number is also called ‘***the square***’ of that number:  
“**3 square**” is 3\*3 or 3² or 9;   
the third power is also called ‘***the cube***’ of the number:  
**“3 cube”** is 3\*3\*3, or 3³ or 27.   
  
One use of **exponential notation**   
is to represent extremely large or extremely small numbers more clearly;  
The number of grains on the 64th cell of the chessboard was “2 to the 64th”,  
a number which would be extremely long to write and difficult to apprehend.  
Better to write it as 264   
  
Because we are used to the decimal system,   
we often represent extremely large numbers in exponential form  
with 10 as the base number.  
Ten billion (10.000.000.000, 10 followed by nine zero’s)   
is more conveniently, written as 109 .  
The universe is about 14.500.000.000 or 14,5\*109  years old,   
whereas mankind appeared some 10\*106 years ago.   
This way of writing makes large numbers more easily comparable.  
In countries with extreme high inflation, money loses its value so fast,  
that the number of zero’s in prices becomes confusing;  
people then start mentioning prices by the number of zero’s:  
“What is the price of this car?” “It is ‘14/8’ Lira” (i.e. 1 400 000 000).

Do you want another example   
of the use of the exponential notation for writing large numbers?  
Take your pocket calculator   
and enter the largest number that it will accept on its screen.  
For me that number is 999 999 999.  
Now add 1 to that number, and you will see the result  
presented as a number in exponential notation  
with base number 10.  
In my case that number is 1e+9, i.e. 1 000 000 000  
or a 1 followed by nine zero’s.  
When your pocket calculator cannot process numbers in the normal notation,  
it switches to the exponential notation.   
You are carrying around exponentials and logarithms (see below)  
in your pocket without being aware of it!

If 2² is 4, 2³ is 8 and 24 is 16, then it follows that 21 is just 2   
while it would be right to write 2 as 21, the custom is to simply write it as ‘2’.  
But there is in fact a ‘silent’ exponent ‘1’ behind any number:   
221, for example, is in fact 2211.   
  
2³ is 2\*2\*2, or if you want 21\*21\*21, hence 2(1+1+1).   
In a similar way (2\*2)\*(2\*2\*2)= (22)\*(2³) = 2(2+3)= 25 = 32.  
  
The rule here is**:   
to multiply two or more exponentiated forms, e.g.** 22 and 2³ **of the same ‘base number’ (2 in this example)   
add the exponents (2 + 3) and raise the base number to that total:** 25**.**For that reason: 100\*109 = 10²\*109 = 10(2+9) = 1011

By the same **rule**, the **exponentiation of an exponentiated number   
is obtained by raising that number to the product of the exponentials**,   
e.g. (2²)³ = 2²\*2²\*2² = 2(2\*3) = 26  
  
One can raise an integer number to a negative integer power and obtain a ‘fraction’.  
We write ½, which is 1/21 as 2-1 ;  
and we write ¼ or (1/2)\*(1/2) or 1/2² as 2-2,  
and we write 1/10 as 10-1, 1/100 as 10-2, and 1/1000 as 10-3.  
  
Extremely small numbers can be expressed more easily or clearly   
in exponentiated form, e.g. .000000001 can be written as 10-9.  
Scientists who talk about extremely small things, e.g. in nanotechnology,  
will not mention the absolute (small) size of their domain,  
but the number of zero’s,   
e.g. ‘ten to the minus nine’ or 10-9 for 0,000000001.  
  
**Multiplying terms with negative powers   
is done by applying the same addition rule** as above,   
e.g.: 3-2\*3-5 = 3(-2-5) = 3-7

and hence 3-2\*35 = 3(-2+5) = 3³

We can now demonstrate an important and useful point,   
namely that 20 = 1;   
indeed: 1 = 2/2 = 2\*2-1 = 21\*2-1 = 2(1-1) or 20 .   
The same holds for any base number: 30 = 80 = 37810 = … = 1   
 **This is an important rule:   
any number, raised to the power zero, is equal to 1**.  
  
Note the following result and rule: 6³ = (2\*3)³ = 2³\*3³;   
when a number (such as 6) can be written as the product of other numbers (such as 2\*3),   
then the exponentiated form of that first number (6³)   
can also be written as the product of those other numbers raised to that exponent (2³\*3³).  
  
The same, of course, holds with division instead of multiplication,   
e.g. 5³ = (10/2)³ = 10³/2³ or 10³\*2-3.  
  
These last rules can sometimes be used to facilitate computations   
e.g. 384/36 = (3\*27)/62 = (3\*27)/(2²\*3²) = 2(7-5)/3²-1 = 25/3.

Notice what happens when exponentiating negative numbers:  
(-1)² = (-1)\*(-1) = 1 (i.e. the square of minus one is plus one);   
This is by the algebra rule that  
**the negative of a negative number is positive**;   
Do you find that strange or hard to understand?   
Just consider this: the reverse (the negative) of a debt is an asset;   
the reverse of being 5000$ in debt is owning 5000$: - (-5000) = + 5000.

Also note that  
(-1)³ = (-1)\*(-1)\*(-1) = -1 (i.e. minus one)  
and (-1)4 = 1; and(-1)5 = -1  
and likewise with -2, and with any other negative number.

We can **write a series of exponentiated numbers in a more general form**;   
instead of writing, for example,   
the series of numbers 1 2 4 6 16 , which is also 20 21 2² 2³ 24,   
we can write more generally 2x  (with x taking on the values 0, 1, 2, 3, 4);   
or, in mathematical writing:

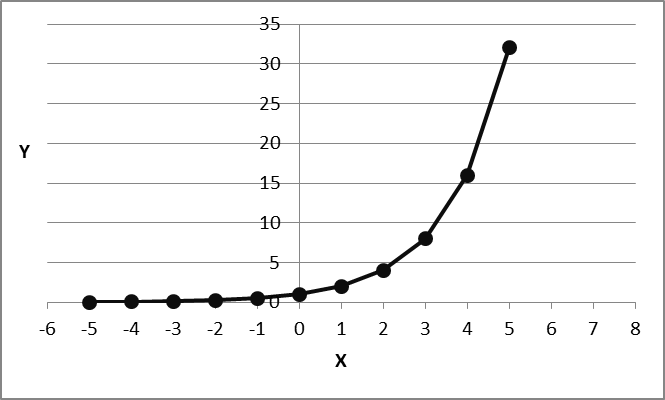
2x (x = 0, 1, 2, 3, 4).

more generally, the infinite geometric series 1, 2, 4, 8, 16,… or 20, 21, 2², 2³, 24, …   
is written in mathematical language as   
 2x (x = 0,1,2,3,4,…);  
  
This series is shown in the table and the corresponding graph below.

x 2x   
0 1  
1 2  
2 4  
3 8  
4 16  
5 32  
… …

The value of 2x can also be computed for negative values of x: -1, -2, -3, etc…

x 2x   
0 1  
-1 ½ or 0,50  
-2 ¼ or 0,25  
-3 1/8 or 0,125  
-4 1/32 or 0,0625  
-5 1/64 or 0,03125  
… …

The upper panel shows that when x becomes large, 2x also becomes (very) large;  
the lower panel shows that when x becomes negative, 2x tends to (i.e. gradually approaches) 0.  
  


The graph and the tables show the value of 2x for integer values of x.  
  
It seems obvious that there must be values of 2x   
that correspond to values of x in-between the integer values shown in the table/graph.   
Stated differently: there must be a value of 2x for each possible value of x,  
also for non-integer values of x;  
and for every value of 2x there must be a corresponding value of x.  
The result is a ‘continuous’ relationship between x and 2x;  
From the picture above, we can ‘sense’ how that relationship is.  
We will return to this below, when discussing ‘functions’   
(2x indeed is ‘a function of’ x).

**3.2 Exponentiation of non-integer numbers.**

Any number can be raised to a power   
including non-integer numbers such as 2,17 or -5,6134, or pi, etc.   
  
For example 2,17² = 2,17\*2,17 = 4,7089   
and 2,17³ = (2,17\*2,17\*2,17);   
a calculator or Excel will be handy to compute how much this is exactly   
(namely 10,218313); but you could just as well leave it written as 2,17³ .  
  
Note that similarly (-2,17)² = (-2,17)\*(-2,17) = 4,7089   
and (-2,17)³ = (-2,17)\*(-2,17)\*(-2,17) = -10,218313.  
  
An example: a container with all three sides equal to 2,17 dm,  
has a capacity of (2,17 dm)³ or 10,218313 dm³ (or 10,218313 liter) exactly.

**3.3 Exponentiating a number to a non-integer power**

As we already hinted above (see the picture of the relationship 2x above)  
if 2² = 4 and 2³ = 8, and since 6 lies between 4 and 8,   
there must be a value for the exponent of 2 (a value ‘x’) such that 2x = 6.   
  
There is indeed a value of x that makes 2x equal 6;   
obviously, that value will lie between 2 and 3 (since 6 lies between 2² = 4 and 2³ = 8)   
and it will be a fractional (non-integer) number.  
  
In the old days, that was not easy to compute   
Today, that number is easily computed   
with an electronic calculator or by Excel (it is 2,584969…).   
  
The conclusion, then is that   
**any number, integer or not integer   
can be raised to any power, integer or not integer**,   
e.g. 2,1780,3334 = 1,296308879 (again computed by Excel…)  
  
Why do you, poor student ‘who does not understand mathematics’,   
need to know and understand all this?  
Well, one important case of raising a number to a fractional power   
is exponentiation to the power 0,50 (or 1/2);   
‘raising’ a number to the power 0,50   
is also called ‘**taking the square root**’ of that number.   
  
For example 20,50  (which you can also write as 21/2)  
is the square root of 2 (which is 1,4142…).   
  
Logically, then, the square of 1,4142, i.e. 1,4142² is 2.   
**Taking the square root of a number is the reverse of squaring that number;**if 20,50 = 1,4142… , then (1,4142…)² = 2.   
  
More generally, if y = x0,50, then y² = x0,50\*x0,50 = x(0,50+0,50) = x(1,00) = x.  
  
Since 4 = 2\*2, but also 4 = (-2)\*(-2) or (-2)²,   
then 41/2 may equal +2 as well as -2,   
this can also be written as 41/2 = +/- 2.  
  
One important implication of this, is that some problems (called ‘quadratic problems’ below)   
that involve taking the square root of a number for their solution  
may have two solutions, a ‘positive’ and a ‘negative’ one.  
For example, below, it will be argued below that firms may have two sizes  
when their profits turn from negative to positive or from positive to negative:  
on the one hand when the firm is too small to be profitable  
and on the other hand when it becomes too large to remain profitable.  
That will be shown to be the result of the ‘quadratic’ nature of the profit function of a firm

Note that there is no solution to the expression (-4)0,50, (or for the square root of any negative number)   
as there is no number, short of an ‘imaginary’ number,   
which multiplied by itself would yield -4 as a result.

The third root (1/3) is called the ‘cubic root’; e.g. 81/3 = 2.  
Note that -81/3 = -2

As we will see below, squaring numbers and taking the square root of numbers   
are operations often performed in applied mathematics and statistics.   
There are good reasons for that, as we show here and later.

**3.4 Sums of squares (and variance and standard deviation of a series of numbers).**  
Let us show why squaring numbers and taking their square root matters.  
  
We return to the series of heights of the eight children.   
We computed their average (or ‘mean’) height to be 111,25 cm.  
  
When we subtract the mean height from each of the original weights,   
*as shown in the Excel sheet 2*we obtain what are called ‘**centered**’ numbers   
(i.e. the same numbers, but now ‘centered’, on the average or mean of the series).  
The centered heights are the differences between   
the height of each individual child and the average height in their group:   
  
 0,75, 3,75, 0,75, -5,25, 8,75, -6,25, -0,25, -2,25.

Positive centered numbers mean that a child is above average in its group,  
negative numbers that it is below the mean of its group.   
  
‘Centering’ numbers or measurements is useful because it tells you   
whether an individual observation lies above or below the average of the group:   
to know that a person is 150 cm does not tell you if it is tall or small  
as long as you do not know the average height of its group.   
150 cm may be tall within a group of 8-year olds,   
but will be small within a group of adult males.  
  
That is what centered values tell us:  
‘is an observation above or below the average, high or low relative to its group?’  
  
But **by how much** is a child in our group of 8 children  
above or below the average of its group?   
Would 5 cm be much?   
And if not, then how many centimeters would we call ‘much’?  
  
We generally have a good feeling for the height of people;   
we know that a difference of 5 cm is not much;   
but when we say such a thing,  
we usually have in mind a reference group or population   
(most likely the adult population), which we do not mention explicitly  
and for which we know that a 5cm difference in height is not much.

But what if these 5cm were about newborn babies?   
Then 5 cm would be a sizeable difference.   
In order to judge whether 5 cm is a large difference or not,  
we must know what are typical differences in height   
in the group that we are talking about.  
  
We must therefore define the ‘typical’ or ‘average’ difference,  
of the heights for this group of children, for this ‘reference’ group   
(or for any group or any other measure that we are dealing with).   
If we know the size of that typical difference,   
we can then, for each individual child,  
(or for whatever it is that we measure, and for whatever reference group)  
say how many ‘typical differences’ it lies above or below the group average.  
Only then can we state something about the child being really tall or small,   
or only slightly taller/smaller than average.  
  
What number would be suited to represent the ‘typical difference’?  
  
We could take the differences themselves, i.e. the centered values  
and compute their average (the mean centered value) in the group.   
Unfortunately, that will not help us   
because the average centered value or average difference is by definition equal to zero:  
the negative and the positive deviations cancel each other out.   
   
 0,75 + 3,75 + 0,75 -5,25 + 8,75 -6,25 -0,25 -2,25 = 0,0   
*as we see in Excel sheet 2;*  
Of course, it would not make sense to say   
that the ‘typical difference’ in the group is 0.  
  
A way to avoid this canceling of positive and negative values is   
to square the differences to avoid the canceling of positive and negative values   
(since squared numbers are always positive)   
then to add up the squared differences to obtain their ‘**sum of squares**’

(0,75)² + (3,75)² + (0,75)² + (-5,25)² + (8,75)² + (-6,25)² + (-0,25)² + (-2,25)² = 163

And finally to compute the average of the eight ‘squared differences’   
[(0,75² + 3,75² + 0,75² + 5,25² + 8,75² + 6,25² + 0,25² + 2,25²)/8] = 20,4375

That number is called **the variance** of the series of numbers;  
the **variance** is a key concept in statistics;   
it is good for you to understand well how it is obtained,  
and to understand that  
a) the larger the differences between the numbers in the series,   
the higher the variance of that series  
b) one large difference may contribute more to the size of the variance   
than several small differences together.  
The square 8.75² (76,56), for example,   
contributes more to the size of the variance  
than 6 other squares together   
0,75² + 3,75² + 0,75² + 5,25² + 0,25² + 2,25² = 47,88.

Just remember this for a while, until we mention the ‘**least squares principle’**  
c) if all measured heights would be equal, the variance would become zero.  
That should be obvious: if all the numbers are equal, there is no ‘variance’ in them  
  
Finally, to obtain a number for the ‘typical difference’,  
a number better comparable to a difference than to a squared difference,   
we ‘un-square’ the variance by taking its square root, i.e.

20,43751/2 = 4,52 cm.  
  
This number, this ‘average deviation’   
is called the ‘**standard deviation**’ (or ‘typical difference’)  
of the series of numbers.   
  
Within our group of children,   
the standard deviation of the heights is 4,52 cm.   
  
If we take any child at random, and ask ourselves   
“by how much we can expect its height to differ from the average in its group?”  
the answer would be “typically by 4.52 cm”.  
  
We see that the first child’s height is less (0,75) than one typical difference   
above the mean height; it is rather average in height.  
The fifth child lies about two ‘standard deviations’ above the mean (8,75); .   
it can safely be called ‘much taller’ relative to its group.   
The sixth child is more than one standard deviation below the mean (-6,25); it is smaller.

In computing the standard deviation, an important statistical number by itself,  
we have illustrated the use   
of squaring values, of summing squares and of taking the square root.

**Summing squared values and variance and standard deviation**  
**are key concepts in statistics**;  
they will be mentioned often in what follows.

But before proceeding to the following topic,   
we introduce one other very important operation and concept,  
the concept of **standardizing a series of numbers or measurements**.

***If you subtract the mean or average from a series of numbers  
you obtain centered values (centered around the average);  
next, if you divide the centered values by their standard deviation,   
you obtain the standardized values of the numbers in the series.***

Standardization means that you place yourself  
in the middle of the observations to measure the heights of the children  
(or whatever else it is that you measure);  
that is what ‘centering’ does,  
and that you measure the size of the differences in height  
using the standard deviation of the series of observations  
as the measurement unit.

*This is shown for the height data in Excel sheet 2:*

0,16 0,78 0,16 -1,09 1,81 -1,30 -,05 -1,09

Standardizing measures is relevant in much research,  
especially in the human and social sciences,  
where many measurements are relative,  
where, as with the heights of children in our example,  
you have to judge their ‘tallness’ or ‘smallness’  
by referring to the average height of their group  
and the typical difference in height for this group.

More generally, standardization of measurements   
is useful when the measurement does not have a clear or natural  
‘anchor point’, nor a clear measurement unit.  
When we count the number of people at a meeting,  
zero (people) is a clear anchor and one (person) is a clear unit.  
When we count 2 persons, we know exactly what that means.

By comparison, however, what does it really mean   
when someone checks the answer ‘2’ on the question  
 “***how much are you in favor of allowing homosexuals to marry?”   
 not at all 1 2 3 4 5 very much***   
  
Does ‘2’ mean that this person really opposes gay marriage?  
(stated differently, does ‘0’ really have any meaning here?,   
or: is ‘3’ really the mid-point between ‘cold’ and ‘warm’?).  
And what does a difference of one point on this scale really mean?  
Is somebody who scores 4 twice as positive as somebody who scores 2?   
  
How ‘absolutely’ can we interpret these numbers?  
Obviously, such numbers cannot be given an absolute interpretation.  
The solution often used is to measure people’s attitudes with such a scale  
and to convert the series of answers to standardized numbers.  
We will return to this many times here and in other courses.

If you find this hard to understand, just consider the following:  
even ‘objective’ numbers like temperatures, e.g. 40°,   
mean something quite different   
depending on whether you use a Celsius or a Farenheit temperature scale.  
40 degrees is very warm in °C, but rather cold in °F.  
(freezing is 0° in °C, but 32° in °F)  
A difference of 1° C is almost 2° in F  
(a one °C difference actually equals a 5/9 °F difference)  
  
If ‘hard numbers’ like those on °C and °F scales   
already differ much in what, e.g., ‘zero degrees’ means   
and differ in what a difference of one degree means,   
then should we not be even more cautious when interpreting   
the meaning of numbers on measurement scales like the one above?

**Standardizing measurements boils down to admitting  
“we do not really have an objective number to represent   
the midpoint or ‘origin’ of this measure,   
neither do we have any objective number to represent   
the ‘measurement unit’ or ‘scale step’ of this measurement”.**  
  
In that case, the commonly used solution is to say:  
“let us place the origin, the ‘zero-point’ of our measurement   
in the middle (the average) of our observations,   
the middle of the group or population studied,  
and let us use the standard deviation of the observations   
in that group or population as our measurement unit”.

That is standardization; often the best you can do.

**Note that when you have standardized a series of numbers  
then, obviously, the mean or average of the standardized values will be 0.  
Less obvious, maybe, is that the standard deviation   
of the standardized values will then be 1,00.***You can verify that for yourself in the example in Excel sheet 2..*  
Standardization is useful to show what you can compare and what not.  
If, for example, we measure more than one aspect of the 8 children,   
e.g. also their weight, *as shown in Excel sheet 7*,  
we see that the mean height is 111,25 and the mean weight is 51,875.   
It would not make sense, of course, to say that these children   
‘on average are more than twice taller than they weigh’ (111,25 versus 51,875)  
Of course not, because these are two different phenomena,  
and you cannot directly compare these numbers.  
Likewise, if we see that the standard deviations are 4,52 cm and 3,76 kg  
can we say that their heights are more dispersed than their weights?  
Again not, of course, for the same reason.  
You cannot compare magnitudes measured on different scales.   
Like it is also the case with °C and °F,  
the height and weight measures have a different ‘anchor’ and ‘step size’.  
  
But if you were to standardize both measures (height and weight),   
then, after standardization, both would be ‘reduced’ to the same status,   
i.e.: **each will have 0.0 as mean   
 and each will have 1.0 as standard deviation.**As a result, after standardization of both measures,  
we will not be tempted to make statements  
that compare heights and weights directly.  
But on the other hand, we might make indirect comparisons.  
Consider child number 5 in Excel sheet 7;   
we may say, after standardizing heights and weights,   
that it lies almost twice the typical difference in height above its group average,  
but only about half the typical difference in weight above its group average:  
relative to its group, the child is much taller than it is heavy.

**Please store this in your long-term memory:  
 after standardization,  
 all measures have the same mean (0.0)  
 and the same standard deviation (1.0)  
i.e. all measures are placed on an ‘equal footing’;  
this allows making comparisons across measurements.**

**Least sum of squares**Now that the concept of a sum of squares is familiar,  
we can introduce the concept of finding the ‘**lowest**’ or ‘**least sum of squares’**.  
  
When explaining how to compute the variance of a series of numbers,  
we saw that this involves summing the squared differences  
between the numbers of the series and their average (111,25):  
  
(112 – 111,25)² + (115 -111,25)² + (112 - 111,25)² + (106 - 111,25)² + (120 - 111,25)² +   
 (105 - 111,25)² + (111 - 111,25)² + (109 - 111,25)²  
  
Let us for a moment pretend that we do not know the value of the mean,  
and ask the following question:  
what is the value of the parameter **m**  
which minimizes the value of the sum of the squared differences   
between the numbers of the series and that unknown number m  
(we do not know m; to find it is the purpose of this question)  
  
This means that we seek the value of m which,  
when entered in the expression  
(112 – m)² + (115 –m)² + (112 - m)² + (106 - m)² + (120 - m)² +   
 (105 - m)² + (111 - m)² + (109 - m)²  
makes this sum of squared differences as small as possible.

Excel sheet 4 allows you to compute   
the value of this sum of squared differences for any value of m  
for example, if m = 115, then the sum of squares is 276  
and if m = 113, it is 188; we seem to be in the good direction!  
If we try 111, it is 164; still better! Let’s continue in that direction!  
If we try 110, it is 176; that is worse! We need to go back…  
If you go on searching like this, you will ultimately find   
that m = 111,25 is **the least squares value (or ‘estimate’) of m**.

With this example, we demonstrate two things:  
- the principle of finding the ‘least squares’ value of an unknown parameter  
by minimizing the sum of squared differences between the observations  
and a number that is, or contains, that parameter  
- the average (or mean) of a series is, in fact, the least squares value of the unknown parameter m  
in this problem.

We will often use the least squares principle  
in order to estimate unknown parameters in what follows.

You will be pleased to know that you do not actually have to seek  
the least squares value of parameters by trial and error  
as we did in this example.  
There are very convenient mathematical formulas or programs  
that allow you to compute these values with extreme speed.  
Some of these are available in Excel.

**3.5 Logarithms**

Logarithms are strange things.  
The concept is so simple   
that you will wonder why they need to be mentioned at all.  
The reason is that they lead to wonderful applications in the real world,   
especially to talk about phenomena that speed up or slow down.  
  
What is a logarithm?  
We saw already that any number (3 in the example below)   
can be represented as some other number,   
called the ‘base number’ or ‘the base’, e.g. 2 in the example   
raised to some power.

For example, with 2 as the base number,   
there is an exponent x which is such that 3 = 2x; namely x = 1,585   
likewise, there is (another) exponent z such that 3 = 2,17z; namely 1,418  
and, of course, there is (still another) exponent w   
such that 3 = 3W (that exponent is 1, remember?).

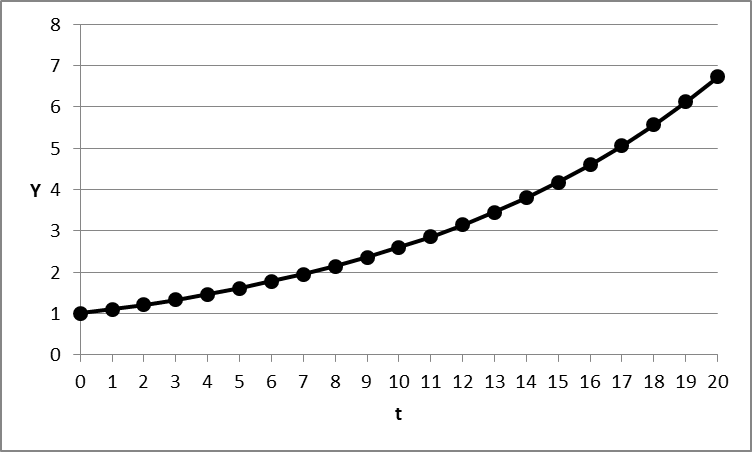
**If x is such that 3 = 2x  
then x is called the ‘logarithm of 3 in base 2’**.  
  
Basically, that’s all there is about logarithms.

Two base numbers are commonly used for logarithmic expressions:  
the natural number ‘e’ (equal to 2,17….) and the number 10.

If 3 = ez (with ‘e’ the Euler number),   
then z is called the ‘***natural logarithm’*** of 3. (z is actually 1,09861)  
The natural logarithm is always written as ln; ln(3) = 1,09861;   
ln() is a function provided by Excel.  
  
Note that, since e is actually e1 , ln(e) must be equal to 1,   
and since e0 = 1, ln(1) = 0.  
  
The logarithm in base 10 is written as log() in Excel;   
as above, log(10) is 1 and log(1) = 0.  
  
To find the natural logarithm of 20, for example,   
we give the instruction = ln(20) to Excel   
and get 2,996 as the answer; so, 20 = e2,996

Likewise, the logarithm of 20 in base 10 is  
log(20) = 1,301  
  
Now, how about this speeding up or down, for which logarithms should be useful?  
Well, we see that the values of e1, e², e³, e4, e5   
are 2,718282; 7,389056; 20,08554; 54,59815; 148,4132  
This shows that in a formula (a ‘function’) such as ex or 2x(or more generally ax, if a is larger than 1)   
when the value of x increases by (only) one,   
then the value of ex   
increases much more rapidly than x.   
If, for example, x doubles from 2 to 4, then 2x, or ex  more than doubles.   
  
While ex increases at an exponential (i.e. increasing) rate,   
the logarithm x increases only ‘linearly’:  
 ln(2,718282) = 1, ln(7,389056) = 2, ln(20,08554) = 3, etc.  
  
That is an important property of logarithms:  
they allow us to relate something that ‘increases with increasing speed’  
to something that increases only at a steady pace.

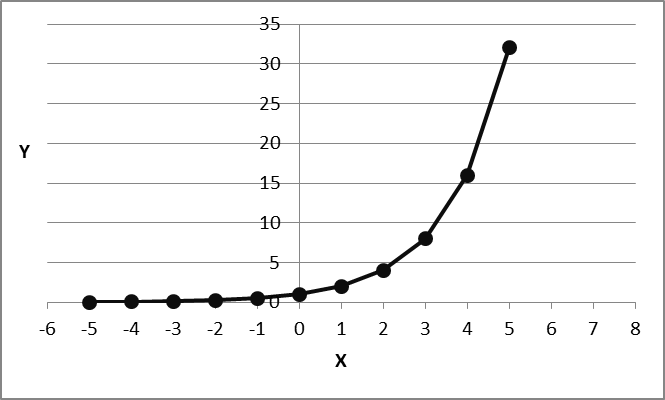
Think, for example, of an object that flies at a steadily increasing speed:   
every minute, its speed increases by 10%.   
That object flies faster and faster,   
but the rate at which its speed increases is a constant 10% per minute.   
While time progresses ‘linearly’, the speed increases exponentially.   
  
Do you find this difficult to understand? Do you want another example?  
Let us look at ‘**compound interest**’ for an example from economics.  
Assume that you deposit 100 dollars in the bank,   
which pays you 10% interest annually on your savings.   
  
At the end of the first year, you receive 10 $ in interest;   
you leave the original sum (100$) and the interest (10$) with the bank, i.e. 110 $.   
At the end of year 2 you receive interest of 10% of 110, i.e. 11 $,   
and your bank account now totals 121 $ (100 + 10 + 11);   
next year the increase is by 12.1 $ and your account now totals 133.1 $,   
and so on...  
  
The result after 5 years is an increase of your savings not by 50%   
(i.e. 100 $ of initial savings and five times 10 $ of interest),   
but by 61,1051%: to a total of 161.061$  
The difference between 50% and 61.1 % is due to the fact   
that you leave the interest with the bank,   
which then ‘pays interest on interest’.   
  
That is why this is called ‘compound interest’;   
the series grows over time (represent time, the number of years by the symbol t)   
as an ‘exponential function’ of t:  
 y = 1,10t (where t = 1, 2, 3, 4, 5)  
 and after 5 years (1,10)5 is 1,61051  
 and after 30 years (1,1)30.is 17,45  
i.e., if you invest 1 $ at 10% interest for 30 years,   
you will receive 17,45 $ for every $ invested!  
  
Saving for pension funds is largely based on this:   
money invested long enough will pay a handsome return.   
Your savings increase in an accelerating way,   
but the rate of acceleration is a constant 10% per year.

  
  
All too abstract?   
Maybe this too will help as one more example:   
Malthus proclaimed that the population of the world  
would increase more than linearly or proportionally with the passage of time.   
Indeed, it would increase at an increasing rate:   
if two humans procreate four children on average,   
who each again procreate four children,   
then the rate of procreation is constant   
but the population grows at an exponential rate,   
the resulting population series will be 4, 16, 64, 256, etc.   
But, said Malthus, the food supply increases only at a constant rate,  
‘linearly’ with the passage of time.  
At some point, that must lead to more people than there is food, to famine and disaster.   
  
This idea (a mathematical model actually!) was quite popular in its time,   
and viewed as an unavoidable law of nature,   
hence as a reason not to do something about poverty and hunger…   
Overpopulation and famine were seen (by the rich and powerful…)   
as unavoidable, ‘a fact of life’, about which nothing could or should be done...  
Such is the power of mathematical models and reasoning over our mind!

At the outset of this text, we said that exponentiation is often used  
to represent and ‘feel’ extremely large or extremely small numbers.   
What we actually do in that case, is mention the logarithm of the number,  
rather than the number itself,   
so that things that grow extremely large or extremely small   
can be represented by a number that increases or decreases less rapidly  
e.g.: 14 000 000 000 = 14\* 109  and 140.000.000.000 is 14\*1010a tenfold increase in absolute value   
corresponds to an increase by only one unit   
in the value of the logarithm of the number in base 10.  
Likewise, 0,000 000 000 1= 10-10 and . 0,000 000 000 01= 10-11

For another example, consider the history of the world since the big bang  
for science, this is 14,5 billion years ago,   
but for the authors of the Genesis, it took only 6 days of ‘creative work’  
(God took a rest on the 7th day…).  
The authors of the Genesis, when talking about the days of creation  
may actually have used something like a logarithmic scale  
to measure the passage of time (the 6 days of creation) from the initial big bang   
to the advent of humans with moral sentiment (Adam and Eve…).

Development (day in genesis) Number of years in the past Ratio to next number

1. Big bang, creation of the universe 15.000.000.000 1,94
2. Stars, our sun and planets form 7.750.000.000 2,06
3. Earth cools, water and dry land appear 3.750.000.000 2,14
4. Earth atmosphere forms 1.750.000.000 2,5
5. First animals appear 700.000.000 2
6. Mammals, hominids and humans appear 350.000.000
7. **Relationships**Relationships have already been introduced above,  
   for example the relationship between x and 2xshown in the picture  
     
   

The graph shows the relationship between 2x and x,  
where the values of 2x are measured on the North-South axis  
(called the **ordinate**)  
and the values of x are measured on the East-West axis   
(called the **abscis)**.  
One typically uses the symbol ‘y’ to represent the values of the function,  
in this case: y = 2x  
and one also usually calls the ordinate ‘**the y-axis**’.  
Likewise, the abscis is more often referred to as ‘**the x-axis’.**  
  
The relationship y = 2x between y and x  
is a complex relationship already.  
The picture shows that it is curved, rather than a simple straight line.  
Let us therefore start with the more simple straight-line relationships

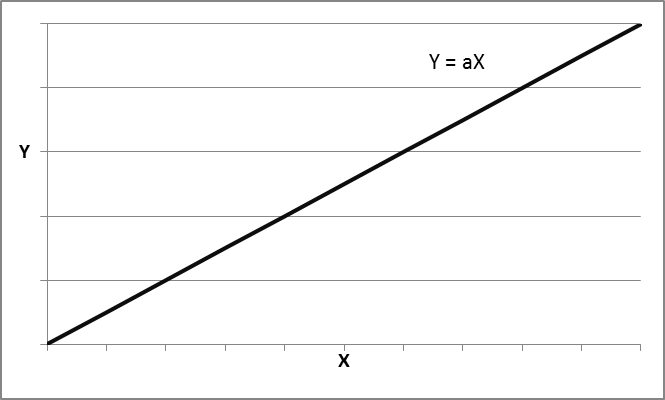
**4.1 Simple linear relationship**

An important phenomenon for us, human beings, is that of ‘relationship’;   
we all understand the word ‘relationship’:   
some systematic association between two or more separate entities;  
for example the relationship between two or more people,  
or the relationship between population density and air pollution,   
or the relationship between driving speed and gasoline consumption, etc.

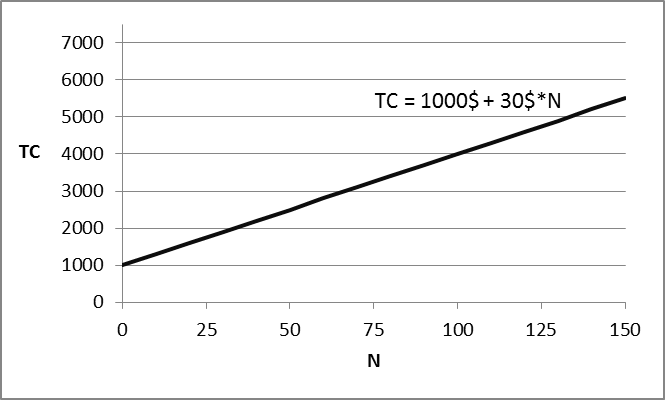
If we understand a relationship,  
we can use that knowledge to predict or manage that phenomenon,  
e.g. the future incidence or respiratory diseases from an increase in urbanization,   
the gasoline consumption of a region if a maximum speed is imposed, etc.

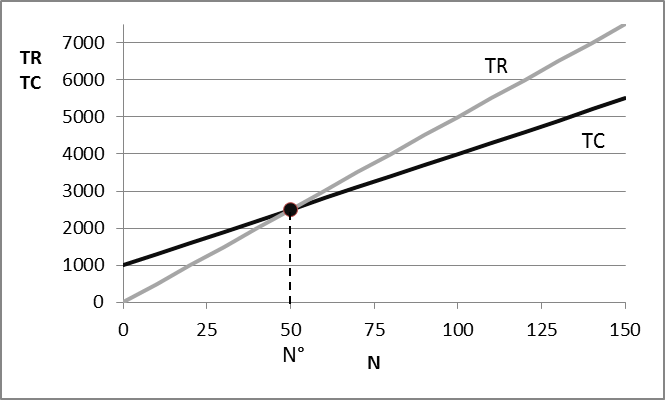
The study of relationships, therefore, is an important part of mathematics  
for those who are interested in doing and applying scientific research.

Let us first explain the concept   
of a relationship between two variables (‘bivariate’).  
  
We use an example from business economics,   
namely a ‘breakeven analysis’.   
**Remember: this is just an example, to demonstrate mathematics in a concrete case;   
this is not about the economics!**  
  
The goal of a breakeven analysis is to determine   
the minimum sales volume that a company needs   
before it starts making a profit.  
  
Take a company that makes Panama Hats   
(Ecuadorians, I know: this is not a hat from Panama,   
but a ‘sombrero de paja toquilla’ from your country!).   
  
Let us say that the company sells its hats for 50$ apiece.   
The revenue (‘TR’) of that company then equals the number of hats it sells,   
represented by the symbol N,   
multiplied by the price it receives for each hat (50$):  
  
 Total Revenue = TR = 50$\*N  
  
This is a ‘relationship’; we can easily represent it as in the graph below.  
  
  
  
A more general expression for a relationship of the kind TR = 50\*N is   
 y = ax

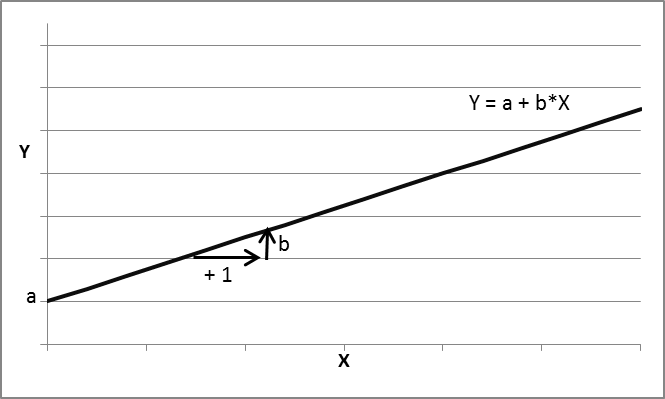
  
here y stands for Total Revenue, x for number of hats , ‘a’ for the revenue per hat (50$);  
y and x are variables.   
Since Total Revenue depends on the number of hats sold,   
TR (or, more generally y) is the **dependent variable**,   
and N (or more generally x) is the **independent variable.**  
a is a ‘**parameter**’  
The parameter a is also called the ‘slope’ of the relationship;  
it expresses by how much y increases if x increases by one unit  
(here, TR increases by 50$ for each increase of sales by one hat)   
if the price changes to 60$ per hat, then ‘a’ is not 50 anymore, but 60.

We can easily compute the total revenue for a given numbers of hats sold,   
irrespective of the price per hat;   
we just need to change 50 into 60 (or any other number) in the formula.   
*Excel is handy for such computations, as demonstrated already in sheet 3.*

Now let us look at the costs of this firm.   
Its costs are of two kinds:   
on the one hand fixed cost (F) for operating a business,   
these costs remain the same, irrespective of the number of hats sold,  
even if you do not sell a single hat   
(e.g. the salary of the manager, the salary of the sales clerk, the rental of the premises);   
on the other hand variable costs per hat (v)   
these are costs for e.g. raw materials and labor that increase with the number of hats.  
If the number of hats doubles, the variable costs double;   
you need double the materials and double the number of hours worked.  
  
Let us say that the fixed cost of the firm are 1000$ and the cost per hat (‘variable’ cost) 30$.  
The total cost (‘TC’) for producing N hats then is, as shown in the graph  
   
 Total Costs = TC = 1000$ + 30$\*N  
  
  
  
or more generally   
  
 Total Costs = F + v\*N   
  
or still more generally   
  
 z = c + d\*x,   
  
where c is the parameter that stands for the fixed costs,   
d is the parameter that stands for the variable cost per unit (per hat),   
z is the variable representing the total costs   
and x is the variable representing the number of hats sold;  
z is the dependent variable, x is the independent variable:   
  
We can now represent both relationships,   
Total Revenue and Total Cost in one graph, as below:



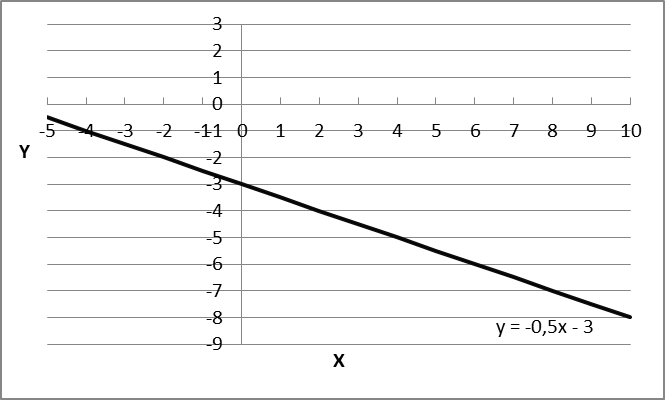
In this graph (called ‘break-even analysis’),   
we see that our hat maker needs to sell a certain number of hats   
before his total revenues are high enough to cover his total costs,   
i.e. before he starts making a profit.  
  
That number, N°, of hats (we write N° to designate one specific value of N)   
where the total revenues exactly equal the total costs   
is called the ‘break-even point’.  
  
As we can see in the graph,   
that is the point where both lines intersect, the point where  
  
 50N = 1000 + 30N  
  
which can be rearranged into (using the rules of algebra mentioned above)  
  
 50N - 30N = (50-30)N = 20N = 1000  
  
from which we can compute N° = 1000/20 = 50 hats  
N° is the solution to the ‘breakeven problem’.

With this example, we have introduced the concept of a line or a ‘linear relationship’:  
both the relationship between number of hats and total revenue and   
that between number of hats and total costs are linear relationships.   
Both are pictured as straight lines.   
  
We are not aware of it,   
but we are using linear relationships all the time  
to explain or picture things to ourselves or to others.  
Like for example the answer to the question  
“When will you be here”?  
“Well, let’s see, I come by bicycle;  
how far is it to your place?”  
“It is about 50 kilometers”  
“Well, I can leave at three o’clock, and I ride some 20 km per hour,  
so I should arrive around 5:30.”  
The reasoning or relationship here is:  
 time of arrival = 3 + (1/20)\*distance, i.e. 3 + (1/20)\*50 = 5:30  
or if you want, y = 3 +(1/20)x   
where y is arrival time, x is the distance.  
   
That is actually a linear relationship,   
if you make a picture of it, it would be a line,  
showing time of arrival ‘in function’ of the distance to be traveled;  
if the distance is 80 km, then time = 3 + (1/20)\*80 = 7.   
  
You will think that this is a stupidly simple example,  
but that is exactly the point:   
the mathematics you read below are in fact very simple.  
  
  
The general expression for a linear relationship is  
  
 y = a + bx,   
  
where y is the dependent variable, x is the independent variable or ‘running’ variable;  
and a and b are parameters (to be given specific values),  
‘a’ is called the ‘**intercept**’;   
that is the value of y when x = 0, as shown graphically below.  
‘b’ is called the slope or inclination of the line;   
it expresses by how much y increases when x increases by one unit.   
If, for example, b = 0.5,   
then y increases by 0.5 each time x increases by one unit (shown in the graph).  
  


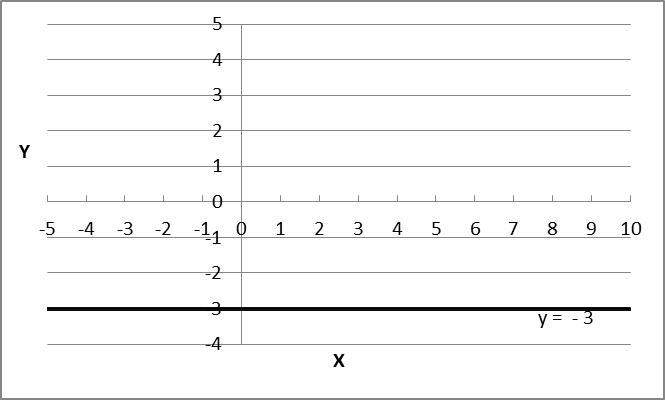
Depending on the value of the parameters a and b,  
you get different relationships between x and y,  
but always linear ones.

To familiarize yourself with linear relationships,  
we show examples y = a + bx   
for positive (+2), zero (0) or negative (-3) values of the parameter a   
and for positive (+4), zero or negative (-0,5) values of the parameter b

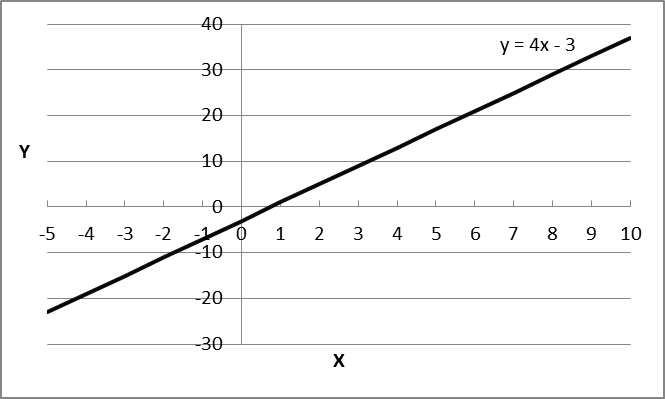
a =-3 and b = -0,5



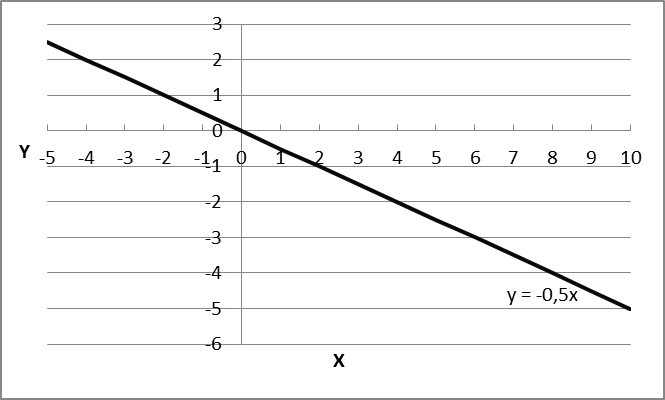
a =-3 and b = 0



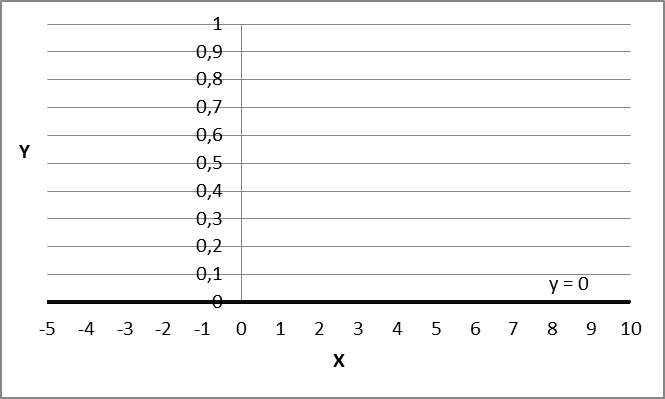
a =-3 and b = 4



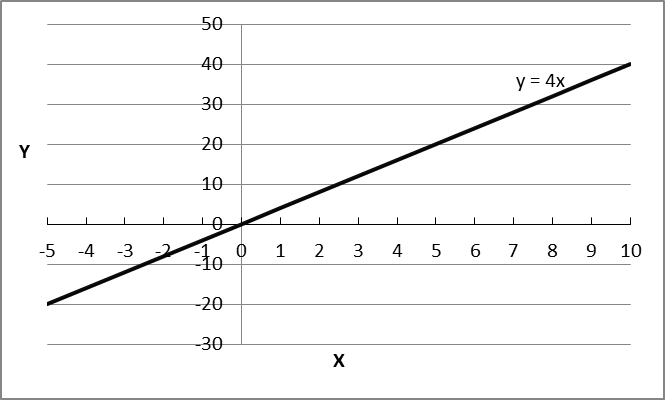
a =0 and b = -0,5



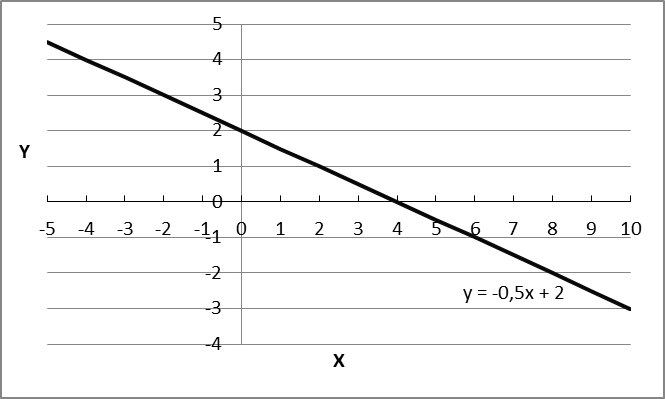
a=0 and b=0



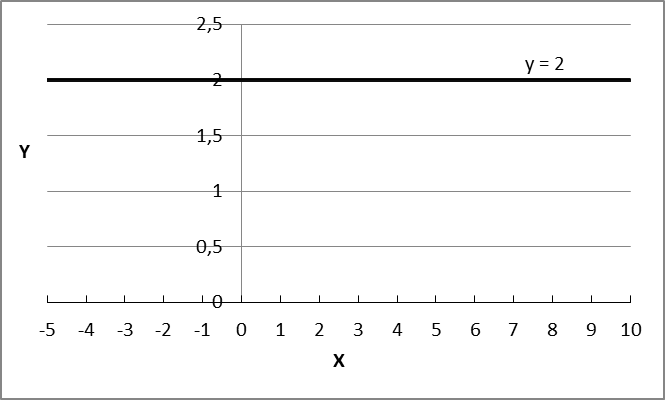
a=0 and b=4



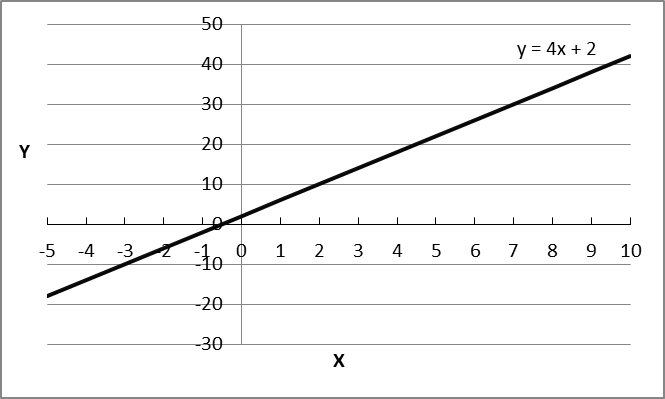
a=2 and b=-0,5



a=2 and b=0



a=2 and b=4



If y = a + bx, then y depends on x:   
if we know the value of the parameters a and b for a specific relationship,   
we can compute exactly the value of y that corresponds with each value of x;  
likewise, since x can be expressed as (y-a)/b or x = -a/b + y/b,   
x can also be computed exactly for any given value of y.

In the case of a linear relationship y = a + bx  
x and y are said to be ‘**linearly dependent’:**if you know the value of one, you know the value of the other.  
In a sense, you might say that if you know one of them (x or y)  
the knowledge of the other is superfluous;   
you can compute one out of the other.   
  
**Linear dependence** will bementioned again below,  
when we do research with several variables  
and it may be that the information contained in one variable is superfluous,  
because it is ‘contained’ in one or more other variables.  
For example:   
if we think that the incidence of cancer in people   
is driven by their extent of smoking   
as well as by their intake of saturated fats,   
but people who smoke also tend to eat a lot of saturated fats,  
it will be difficult to find out how much of the incidence of cancer  
is due to smoking and how much to unhealthy eating  
Smoking and unhealthy eating then tend to go ‘hand in hand’,  
tend to be **linearly related** to each other.  
It then becomes difficult to separate out the effect on cancer  
of smoking on the one hand and unhealthy eating on the other.  
But that is for later discussion…

Now, back to linear relationships.  
  
As we all know, if you take a page of paper and plot two points on it,  
you can draw exactly one, and only one, straight line through these two points.  
Any two points (x, y) suffice to define a line or linear relationship.  
If we know that a line runs through the points (x = 2, y = 0) and (x =5, y = 6),   
since the line (the linear relationship) is of the general form y = a +bx,  
then the following must hold  
 0 = a + 2b (y = 0 when x = 2)   
 6 = a + 5b (y = 6 when x = 5)   
  
Subtracting the first from the second relationship,   
we obtain 6-0 = a-a + 5b-2b   
 or 6 = 3b, and hence b = 2  
  
substituting b = 2 in the first expression, (0 = a + 2b)   
we find that 0 = a +2\*2 and hence a = -4.  
  
The expression of the relationship y = a +bx through (x = 2, y = 0) and (x =5, y = 6)  
therefore is y = -4 +2x  
  
We can verify that (x = 2, y = 0) and (x =5, y = 6) meet the condition:  
 indeed: 0 = a + 2b or 0 = -4 + 2\*2 (for the point 2,0)   
 and 6 = a + 5b or 6 = -4 + 5\*2 (for the point 5,6)

Once you know the parameters a and b for a relationship of the form y = a + bx,   
you can use it to compute any value of y within the original range of values of x,  
i.e. between 2 and 5, for example x = 3; that is called ‘**interpolation**’;   
 e.g. if x = 3, then y = -4 +2\*3 = 2  
  
You can also use it to compute a value of y   
for a value of x outside of its original interval,   
i.e. outside the range from 2 to 5,   
 e.g. if x = 1000, then y = -4 +2\*1000 or 1996.   
The latter is called **‘extrapolation’**.   
  
When interpolating, you are ‘in known territory’;   
when extrapolating, you are ‘in new territory’.  
Scientists will often develop a model based on observations of past conditions   
(i.e. interpolation)  
and use that model to predict what will happen under new conditions  
(i.e. extrapolation)

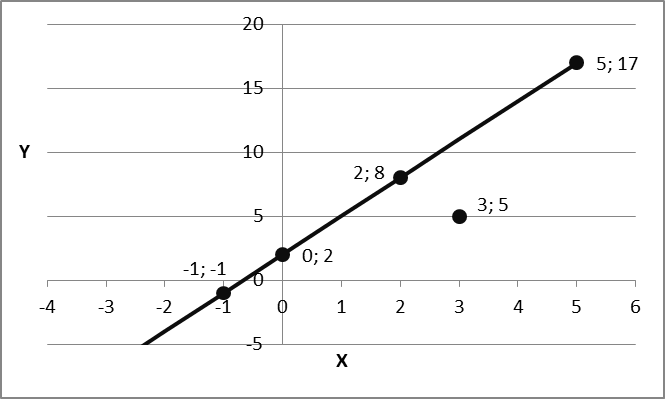
If two lines y = a+bx and z = c+dx intersect   
(i.e. y and z have the same value),   
like that was the case above in the break-even analysis   
with the Total Revenues line and the Total Costs,   
that will, by definition, be for a specific value of x, call it x°, such that y = z,  
  
 i.e. a+bx° = c+dx°   
 or, using the rules of algebra bx° - dx° = c – a or (b-d)x° = (c-a)  
 and hence x° = (c-a)/(b-d)  
  
For example, in the example with hats, where y = 50x and z = 1000 + 30x   
i.e. with a = 0, b = 50, c = 1000 and d = 30   
that is when x° = 1000/(50 – 30) = 50 hats, as already shown above.  
  
The expression x° = (c-a)/(b-d) illustrates the uses and advantages of mathematics:   
it is a general solution to a problem (‘where do two lines intersect?’),   
which can then be tailored to any specific situation   
(‘when do the Total Revenues of a hat manufacturer equal the Total Costs?)  
i.e., to specific values of the parameters a, b, c, d;   
this avoids having to compute a specific solution   
every time you have a problem of the same structure  
but with different parameters   
(e.g. the intersection of two lines)   
  
If the price per hat drops to 40\*, for example,   
the break-even volume becomes 1000/(40-30) = 100.  
  
If a =10, b = 100, c = 2000 and d = 20,   
then x° can be computed with the same rule as (c-a)/(b-d) = (2000-10)/(100-20)

**x° = (c-a)/(b-d), again, is a rule to remember.  
You do not have to prove it again, the proof has been given;  
you can now use it for any problem where two linear relationships intersect.**

As we saw in the graph above, the relationships  
   
 TR = 50\*N and TC = 1000 + 30\*N   
  
are lines, or linear relationships.   
This means that if we sell one more hat,   
the total revenue increases by a fixed 50$   
(irrespective of whether we already sell few or many hats),   
or if we manufacture one more hat, the costs increase by 30$   
(again, irrespective of how many hats we are already manufacturing at that point).  
  
These relationships can be visualized on a sheet of paper,   
i.e. on a flat map, a plane,   
with an axis for TR or TC (the up-down axis, or North-South axis, usually identified as y)   
and an axis for N (the left-to-right axis, or East-West axis, usually identified as x).   
  
Each point of that plane has a coordinate for N and for TR (or TC)   
or, more generally, for x and for y.  
  
A line, then, is the collection of points (x,y) in a map or plane  
that satisfy the condition y = a + bx for some specific values of a and b.  
For example, the line TC = 1000 +30N   
is the collection of points with coordinates (N, TC)  
such that TC = 1000 + 30N  
Evidently, 1030 lies on that line (it is the value of TC for N =1)   
and so does 1300 (the value of TC for n = 10);  
but the point (N, TC) = (10, 2000) obviously does not lie on the line.  
  
Similarly, the line y = 2 + 3x   
is the collection of all points (x°,y°) in the plane with axes x and y   
which meet the requirement that

y° = 2 + 3x°

The point (2,8) is such a point, since 8 = 2+3\*2; so is the point (5, 17);   
and so are also the points (0, 2) and (-1, -1);   
on the other hand, the point (3, 5) does not lie on this line,   
since it is not true that 5 = 2 + 3\*3.

This is illustrated graphically below.  
  


Why did we need to mention this?  
Below, we will see that in reality,   
the observations (points) that we can make of a relationship  
do not always lie neatly on a straight line, but almost.  
We will then see how we can nevertheless ‘guess’  
what the linear relationship is below  
in the section on estimating the parameters of a linear relationship.

**4.2 Multivariate linear relationships.**

In our example, total revenue depends only on the number of hats sold;  
the relationship is between two variables, TR and N (or y and x);   
it is **bivariate**.  
  
Linear relationships can involve more than two phenomena or variables.  
Let us illustrate that by expanding the Panama hat example.   
If the firm also makes ties, and sells them at 20$ apiece,  
then its total revenue depends not only on the number of hats sold (N),   
but also on the number ties sold (T):  
  
 Total Revenue = 50\*N + 20\*T  
  
or in general terms:  
 y = a + bx + cz   
  
(with parameters a, b and c; in the example a = 0, b = 50 and c = 20).  
  
If we want to represent this graphically,   
we are not looking at lines on a map (i.e. in a flat ‘plane’) anymore,   
but rather at a plane in a space with axes TR, N and T;  
(we are ‘moving up’ one dimension in our reasoning:   
the unidimensional line of the previous example   
now becomes a two-dimensional plane   
and the 2-dimensional plane in which we drew te line  
now becomes a 3-dimensional space, i.e. ‘space’ as we are used to think of it,  
with a right-left, front-back and up-down dimension or axis).   
  
Each point of that plane meets the condition that TR = 50\*N + 20\*T.   
  
That plane goes through the origin   
(TR = 0 when N = 0 and T = 0; i.e. through the point 0, 0, 0),  
this expresses that the revenues are zero  
when the firm does not sell any hat or tie.  
The plane slopes upwards with N and with T,   
more steeply (50 per unit) with N than with T (30 per unit).   
  
The relationship between Total revenue y), hats sold (x) and ties sold (z)   
is illustrated in the picture below (a graphical model…).

While TR = 50$\*N or y = a + bx is a **bivariate** linear relationship,   
TR + 50\*N + 20\*T or y = a + bx + cz is a **multivariate** linear relationship   
(in this case it is **trivariate**, between three variables: y, x and z).

Such a trivariate relationships may be useful to describe, and build,   
a sloping flat surface, e.g. in architecture or in design.   
But it is also useful to represent or model phenomena   
where one dependent variable is caused   
by more than one independent variable,   
such as when the revenues of a firm depend on N and T,   
or when the incidence of cancer depends   
on the extent of smoking and on the intake of saturated fats.

If the firm also sells S shirts, at 80$ per shirt,   
then the linear relationship would become TR = 50N + 20T + 80S,   
and so on…

**A final note on linear relationships: linear transformation.**

To illustrate what we mean by a **linear transformation**,  
let us return to the example of measuring temperature.  
Temperature remains the same,  
irrespective of whether we measure it in °C or in °F.  
It must therefore be possible to ‘translate’ or ‘transform’ °C into °F and vice versa.  
  
Since we know that water freezes at 32°F, or at 0°C  
and since a 1° difference in °F is a 5/9° difference in °C,  
the transformation is given by °C = (°F-32)\*(5/9)   
or °C = – 32\*(5/9) + (5/9)\*°F, this is a linear relationship of the form °C = a \* b\*°F  
for example, 65°F corresponds to (65-32)\*(5/9) °C = 18,33°C  
Using the rules of algebra, we can also say that  
°C\*(9/5) = °F – 32, or °F = 32 + °C\*(9/5),   
again a linear relationship, of the form °F = c + d\*°C  
  
Another example of a linear transformation is given   
by the operation of standardizing a measurement  
The standardized value z of a measurement x is obtained as  
 z = (x-m)/s  
with ‘m’ the mean or average of the measurements and ‘s’ their standard deviation.  
Using the rules of algebra, this can be written as z = -(m/s) + x/s  
if we substitute the symbol ‘a’ for –(m/s) and ‘b’ for (1/s)  
then we can write z = a + bx, a relationship with linear form.

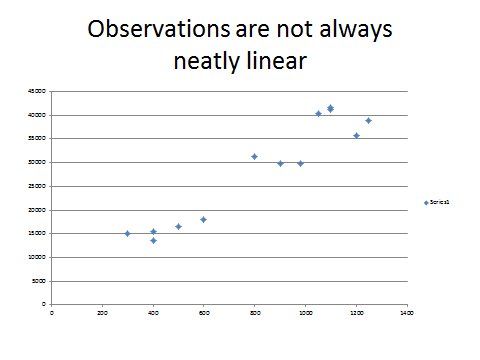
**4.3 Estimating the parameters of a linear relationship**

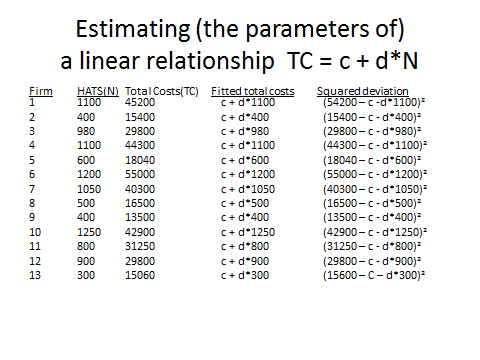
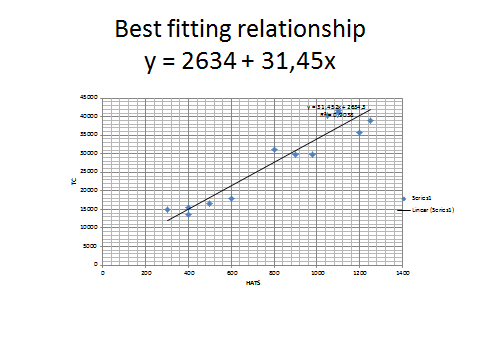
Typically, in scientific research, we do not know exactly the form or model  
taken on by a specific phenomenon, e.g. the relationship between two variables,  
such as the incidence of lung cancer and air pollution;   
we may assume that it is linear, e.g. of the form y = a + bx,  
but we usually do not know the value of the parameters a and b.

We may have to guess the value of these parameters ourselves,   
or try to infer their value from the past history of the phenomenon.  
  
To use a simple and familiar example,   
let us revert to the hat manufacturer.

When the manager of the hat company says   
that his Total Revenue is 50$\*N.   
that may actually only be his assumption or guess.   
Is this really true? Do the sales figures confirm this?  
Sometimes, hats may be sold at a lower price,   
e.g. when the customer buys more than one,  
or sometimes rich customers may be charged a higher price…  
Then, what really is the relationship between revenue and number of hats?  
Can we infer the parameter of the relationship TR = a\*N  
from past observations on number of hats and revenues?  
  
And what is the relationship between Total Costs and number of hats;   
is it really 1000$ + 30$\*N as the manager says?  
Can we infer the parameters of the cost function TC = F + v\*N  
from inspecting past results of hat manufacturers,   
from the number of hats produced last year and total costs in that year?  
  
Assume that the Hat makers’ association provides us with the Total Costs   
and the number of manufactured hats for 13 firms, as in the table below.

Firm HATS(N) Total Costs(TC)  
1 1100 45200  
2 400 15400   
3 980 29800   
4 1100 44300   
5 600 18040   
6 1200 55000   
7 1050 40300   
8 500 16500   
9 400 13500   
10 1250 42900   
11 800 31250   
12 900 29800   
13 300 15060

If we make a graph, as below, then we see that the observations, the data   
do not all lie neatly on one straight line TC= c + dN;   
it is not the case that TC is exactly equal to 1000 + 30\*N for each firm.   


Then what is the (linear) relationship between total cost and number of hats   
 TC = c + d\*N   
which best reflects the reality (the data)?;   
what is the ‘best’ line TC = c + d\*N.  
That is to say: what is the best value for the parameters, c° and d°,   
so that the relationship TC = c° + d°\*N closely approximates the observations in the graph?  
  
If I knew these ‘best’ values of ‘c’ and ‘d’ (call these c° and d°),  
then for Firm 1, the relationship would predict Total Costs of c° + d°\*1100;   
since the real costs for that firm are 45200;   
the difference between reality and the prediction is (45200- c° – d°\*1100);   
for Firm 2 the difference would be (15400 -c° - d°\*400), etc.  
  
We look for values for c° and d° that make these differences small.   
  
But we do not know c° and d°;   
the best we can do is estimate them on the basis of the available data.  
To find, or ‘estimate’ c° and d°, we apply   
the ‘**least sum of squared differences**’ or ‘**least squares**’ principle  
that was introduced above:   
look for the value of ‘c’ and ‘d’ which make the sum of the 13 squared differences   
(45200- c – d\*1100)² + (15400 -c - d\*400)² + … + (15600 – c – d\*300)²   
as small as possible.   
  
For these values of ‘c’ and ‘d’ the resulting ‘fitted’ or predicted Total Costs   
will lie ‘as close as possible’ to the real Total Cost figures;  
the squared difference between reality and the prediction will then be the smallest   
Since there are 13 squared differences (one for each month),   
we sum them, add them up and look (we ask the computer to look for this…)  
for the value of ‘c’ and ‘d’ that make that ‘sum of squares’ minimal.  
  
That is an operation easily performed in Excel;  
e.g. by the function LINEST in Excel under Formulas>more functions>statistics>linest  
or even as a facility offered together with the scatter plot of the data..  


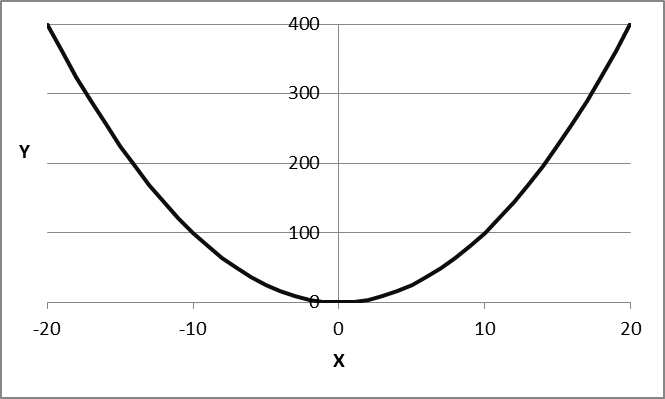
The estimated relationship y = a + bx, in this case TC = c + d\*N  
is seen to be   
 y = 2634 + 31,45\*N  
It is called the ‘**simple linear regression of y on x’**,   
here it is the regression of TC on N.  
  
We say ‘linear’ because the relationship is of the form ‘ y = a + bx’  
and we say ‘simple’ because this involves only two variables, X and y.  
Simple linear regressions are used very often in empirical research to model  
relationships between one dependent variable (y) and one independent variable (x)  
and where there is no reason to think that the relationship is not linear (see below).   
  
If y is assumed to depend on more than one variable, for example  
  
 y = a + bx + cz

then the best fitting linear relationship will be a ‘multiple linear regression’.

**4.4 Nonlinear relationships: quadratic relationships.**  
Up to now we discussed linear relationships:   
i.e. relationships y = a + bx, or actually y1 = a + bx1where the variables are just in ‘pure’ form.

In such a relationship, when x increases by one unit,   
y always increases by b units,  
irrespective of whether x increases from 1 to 2 or from 1000 to 1001,  
irrespective of whether we are operating at lower or higher ranges of x.  
  
The latter will not be the case in nonlinear relationships,   
there, the impact on y of an increase of x by 1 will differ  
depending on whether x is already high in value or not.  
  
A simple example of a nonlinear relationship is the **quadratic** relationship (or ‘function’)  
  
 y = x²

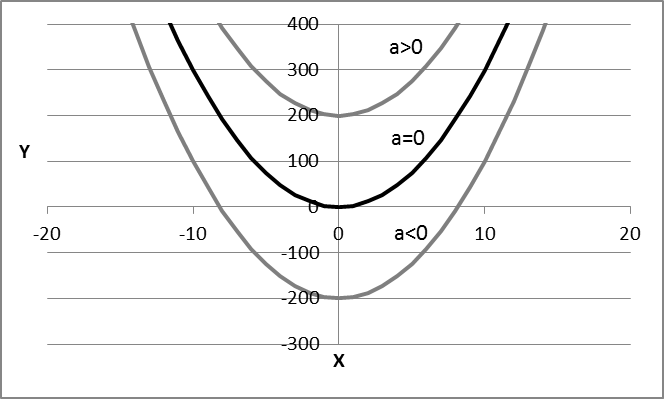
This relationship is shown in the graph below.   
  
It is symmetric around the value x = 0;   
it increases at an increasing (i.e. nonlinear) speed as x moves away from zero.  
It illustrates the concept of an **extremum**, of a function with an extreme value;   
in this case the extremum is a **minimum** that occurs when x = 0   
since our model is y = x², the minimum of y is (also) 0.

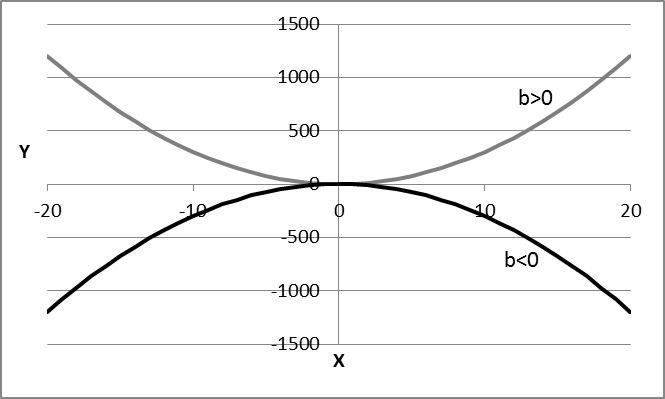
  
  
  
If we want to move that relationship up or down, we may write   
 y = a + x² (up for a > 0; down for a < 0);   
now the minimum is not y = 0 anymore, but y = a

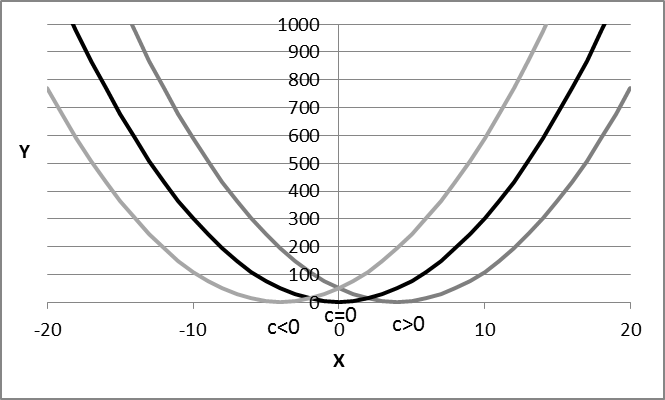
If we want to move the relationship to the right or the left, we may write  
 y = (x-b)² (to the left if b < 0; to the right if b> 0);  
now the minimum occurs not when x = 0, but when x = b.  
   
For a quadratic relationship with a maximum, rather than a minimum, we write  
 y = c + ax² with a < 0

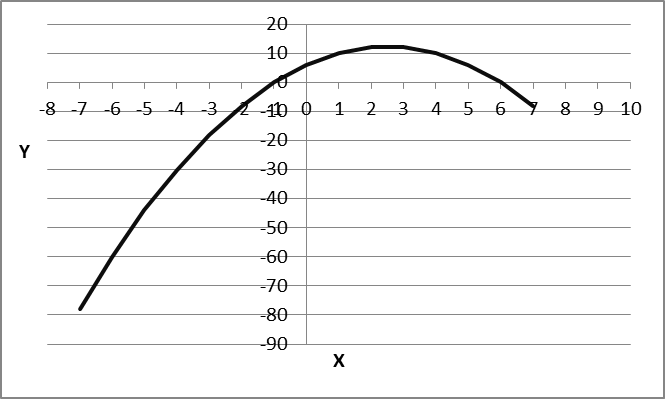
If we want to make the relationship broader or tighter, we write  
 y = ax² (broader for larger values of ‘a’; tighter for smaller values).

More generally then, we can write a quadratic relationship as  
 y = a + b(x-c)²  
With a, b and c as parameters.   
The parameter ‘a’ shifts the relationship up or down,  
the parameter ‘c’ shifts it right or left  
the parameter b makes it broader or tighter.  
  
We show pictures of this relationship for various parameter values   
a (> 0, = 0, < 0), b (> 0, < 0) and c (> 0, = 0, < 0).







A still more general form of a quadratic relationship between y and x   
is given by the general expression  
   
 y = a + bx + cx² (where a, b and c are parameters)  
  
for example y = 6 + 5x – x2  
  
With an Excel sheet it is easy to compute  
the value y of such a function for successive values of x;   
and examine   
 - where (i.e. for what value of x) the function reaches an extremum   
(a maximum or a minimum),   
- whether (and if so, where, for what value of x) the function turns   
from positive to negative or from negative to positive,   
(such values of x are called the ‘**roots**’ of a function).  
   


We do not give the proof , but mention that  
**- a function of the type y = a + bx + cx² has an extremum when x = -b/2c  
- if a function of this type has a value of y = 0 for some value of x,  
i.e. if there is a value of x where y turns from positive to negative  
or from positive to negative   
that will be when x has the value   
 {-b + SQRT(b²-4ac)}/(2c) or {-b - SQRT(b²-4ac)}/(2c)**The latter are called the ‘**roots**’ of the quadratic equation,   
i.e. the value(s) – if any - of x for which y is zero.  
  
The rules just mentioned are not so simple;   
but it pays to learn them by heart, to become habituated to them.

Note that a quadratic equation must not necessarily have roots  
i.e. values of x where the function changes sign)  
for example, the quadratic relationship y = 4 + x²   
is a relationship with a minimum value of 4 (when x = 0);   
its value is never less than 4;   
it never changes from positive to negative;   
hence it has no roots.

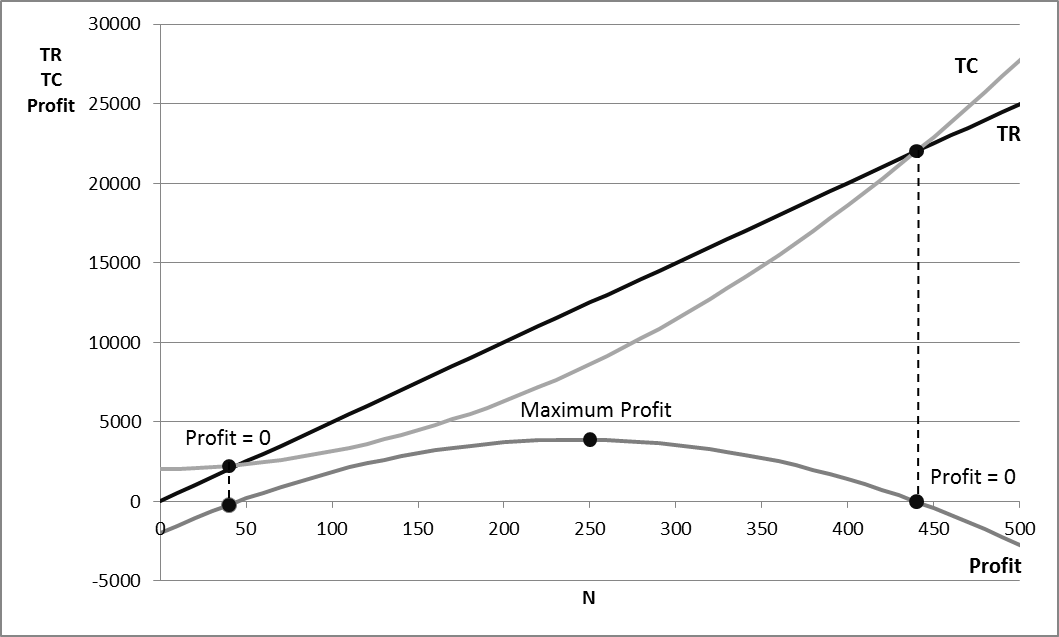
In the case of the example y = 6 + 5x – x2,   
the extremum (a maximum) occurs when x = -5/(-2) = 2.5   
and the roots are x = -6 or or x = +1, i.e.   
 (-5 + SQRT{25- 4\*6\*(-1)})/(-2) or (-5 - SQRT{25- 4\*6\*(-1)})/(-2)

Can we give an example of such a relationship from ‘real life’?   
Just think of shooting an arrow from a distance at a target;  
when we shoot an arrow, we intuitively take into account   
that it will sag while flying;   
to correct for this, we aim slightly above the target,  
because we know intuitively   
that the flight of the arrow first rises to a maximum and then declines.  
What we do intuitively,   
is let our arrow follow a quadratic curve with a maximum..  
  
If our intuition is not enough,   
there is a military science called ‘**ballistics**’, devoted specifically   
to computing curvilinear (quadratic) trajectories for objects (e.g. bombs)  
so that they reach their target.  
  
Do you want another example of a quadratic relationship?   
Here is one from economics   
(remember: it is not the economics, but the mathematics that interest us here).   
  
Take the hat-producing firm described above;   
we modeled its total costs as y = F +v\*N  
This implies that, as the firm produces more hats (i.e. as N increases),   
its costs simply increase linearly (i.e. in line with N1).  
  
In reality, when firms raise their production volume by a large amount,   
their costs will start to increase more than linearly with N:   
more, and less efficient workers must be hired,  
the larger number of workers will require an added level of supervision,   
communication and coordination problems will increase, etc.   
  
The result is that the total costs (TC) of the firm   
start to increase sharply beyond some production level,   
more than proportionally with N,   
e.g. proportionally to N²:

Total Costs = F + v\*n + q\*N², e.g. TC = 5000 + 30\*N + 0,1\*N²

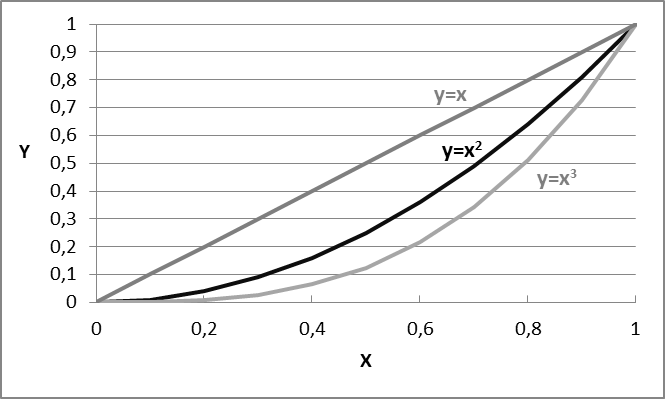
where q or 0,1 is the parameter expressing the ‘quadratic’ incidence  
(more than proportional, more than linear)   
of the production volume N on Total Costs.  
  
TC now is a quadratic function of N.   
It has the general form y = a + bx + cx2   
When N becomes large, N² becomes even larger,   
and Total Costs start increasing sharply.

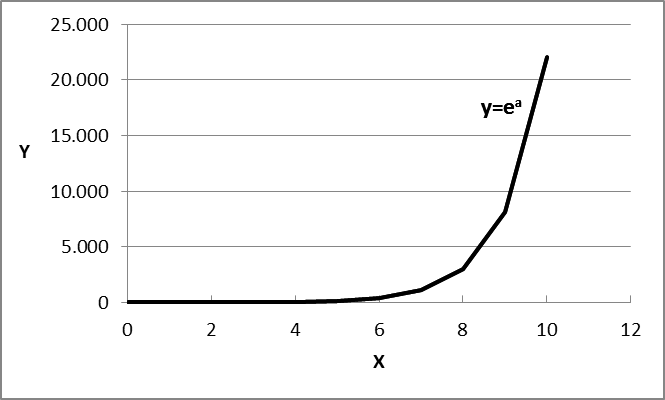
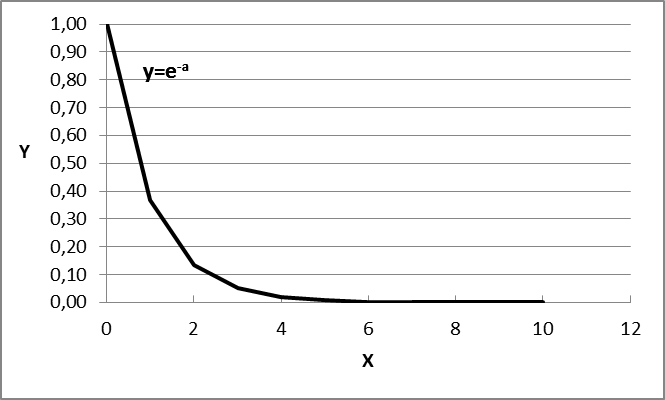
The profits of the firm are its Total Revenues (TR = 50$\*N, as above)   
minus the Total Costs (now TC = 5000 + 30\*N + 0,1\*N²).  
  
Since the total revenues and the total costs   
depend on the number of hats produced (N),   
the profit (TR - TC) also becomes a function of N,  
or more correctly a function of N and of N².  
  
 Profit = TR – TC = a\*N – (F + v\*n + q\*N²) =  
 = 50\*N – (1000 + 30\*N + 0,1\*N²)  
 = -1000 + (50 – 30)N – 0.1\*N²  
  
The profits have now become a quadratic function of the production volume N.  
The profit function will therefore have an extremum, an optimum  
and roots (where profits turn into losses).  
We can approach the problem of the firm as one of optimization:  
find the optimal level of N, the optimal production level.

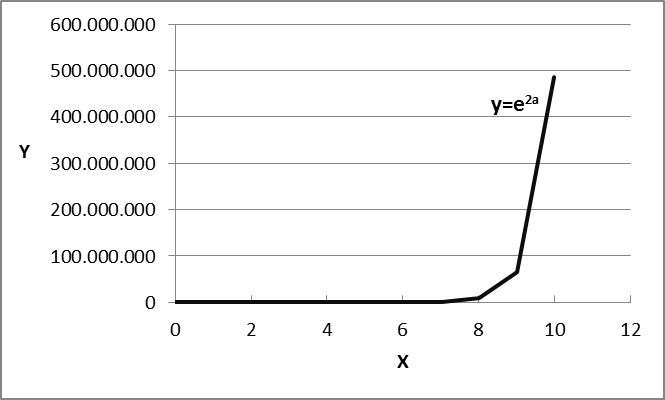
**Optimization.**Let us revisit the ‘break-even’ problem discussed above.   
There, the costs were assumed to be a linear function of N;  
in that case there is a (breakeven) point N° of production and sales   
above which the firm becomes profitable;   
from there on, the more hats it sells, the more profit it makes.   
  
With an infinite number of hats sold, profits become infinite;  
that is not a very realistic assumption;   
it is the result of using a linear models to represent revenues and costs,  
so that each additional hat sold brings in the same amount of additional profit,  
irrespective of whether we are currently selling few or many hats.   
  
With a quadratic cost function, however,   
beyond a certain production volume,   
the costs will start to increase so much,   
that the total costs will exceed the total revenues by a large amount.   
  
First, at low sales volumes,   
revenues will be too small to cover the fixed costs; the firm then makes a loss;   
then, as the production volume N increases,   
there is a point where the firm breaks even and starts making a profit;  
as N increases further, profits will grow and peak; they reach a maximum;   
next they will start to decline again;   
and ultimately, at very high production volumes,   
the firm will run up such high costs that it will be making losses again.  
  
That is represented graphically below in the profit curve:   
it is clearly a nonlinear relationship, with an extremum and two roots.   
The extremum is where profits are maximal;   
the roots are where profits change sign (from profit to loss)  
  


With this example, we also illustrate the concept of **optimization**.  
In many cases, our model of a phenomenon   
allows us to ask and answer the question  
“how we can bring the phenomenon to a desirable or optimal value?”.  
In the example above, the question is how the firm can maximize its profit:  
what is the number of hats, N°, where the profits are highest?  
  
Using the ‘theory’ above, we know that this will be the value of N  
where x = – b/(2c) or N° = -(50-30)+(-2\*0,1) = -20/(-0,2) = 100 hats  
  
This example is one of **static optimization**.  
The decision to be made is only the number N of hats to produce.  
Static optimization problems differ from dynamic optimization problems.  
  
**Dynamic optimization problems**   
  
For their optimization, these problems ask   
for more than just a single value of the decision variable;  
rather, they ask for an optimal ‘**trajectory**’ of the decision variable over time.  
For example, landing a module on the planet Mars   
requires more than just one value of the thrust of the rockets   
to propel and slow down the module,  
(that would be a static optimization problem);  
rather, what is required is a time-path or ‘trajectory’ of the decision variable   
defining the needed thrust at each moment  
from the time of launching the module   
until the time of its landing on Mars.  
This is also called an ‘optimal control problem’;  
such problems occur in many walks of life.  
  
Let us point out for now, that many efforts in scientific research  
deal with setting up models that require optimization,  
and with (procedures for ) finding the optimal value of the decision variables.  
  
If, for example, we have a model of the spreading of an infectious disease,  
and of the impact of vaccination on the contraction of that disease,  
an optimization problem might be to find out   
the optimal intensity over time of vaccination efforts,  
and to answer questions such as   
“is it better to vaccinate very early in the diffusion of the disease,  
or rather at the moment when the spread of the disease speed up?”  
(see below for models of diffusion).

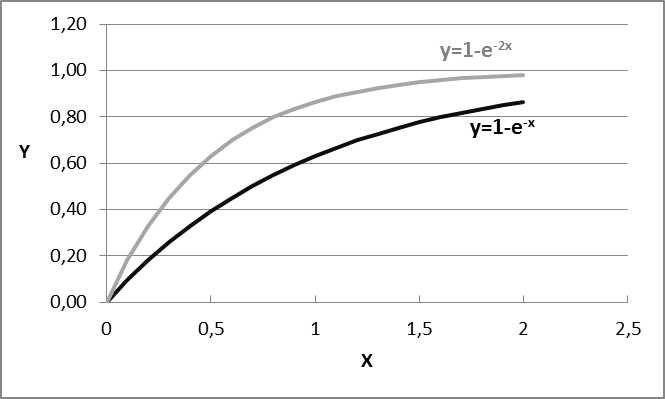
**4.5 Other nonlinear relationships.**  
  
There are an endless number of types of nonlinear relationships   
next to the quadratic relationship discussed at length above.   
  
We limit ourselves to some that are often used in scientific research  
(the choice may reflect the discipline of the author: economics and psychology):  
  
- relationships that relate y to a power of x: y = xa  
 e.g. the **square root function** y = x1/2 (SQRT(x) in Excel notation);   
  
Here is an interesting application of the model y = xa:  
if you arrange the population of a country from poorest to richest  
and for each % of the population compute   
what % of the wealth of the country they own,  
i.e. what % of the wealth do the poorest 1%, 2%, 5%, 10%, etc., possess?  
until you have counted everybody   
(and 100% of the people own 100% of the wealth),  
you obtain curves of the kind y = xa such as pictured below   
(we show such a curve xa for a = 1, 2 and 3).  
  
The closer the curve lies to the diagonal, the more egalitarian that society,  
(egalitarian means that the ‘poorest’ x% own x% of the wealth);  
the further below the diagonal this curve, the more unequal the society,  
i.e. then the poorest x% own much less than x% of the wealth.  
The parameter ‘a’ in the relationship xa can therefore be used   
as an index of economic (in)equality  
and used to compare the economic inequality between countries.  
This (the parameter a) is called the Gini-coefficient in economics.

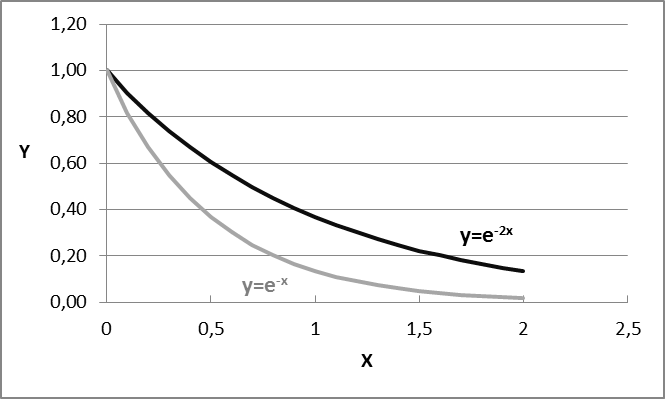
, 

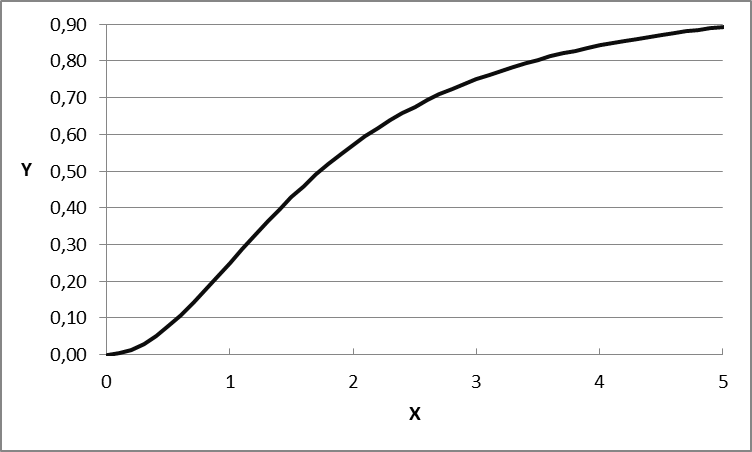
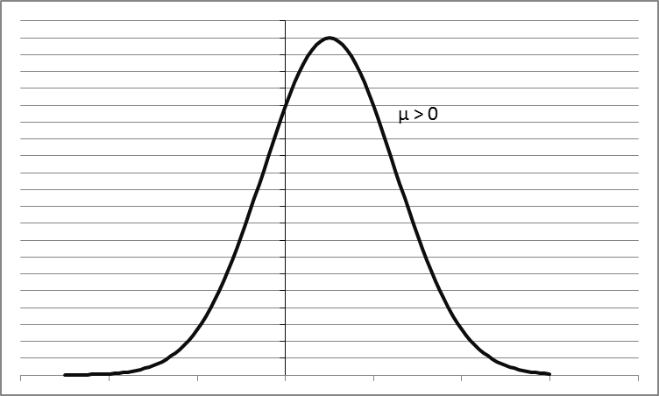
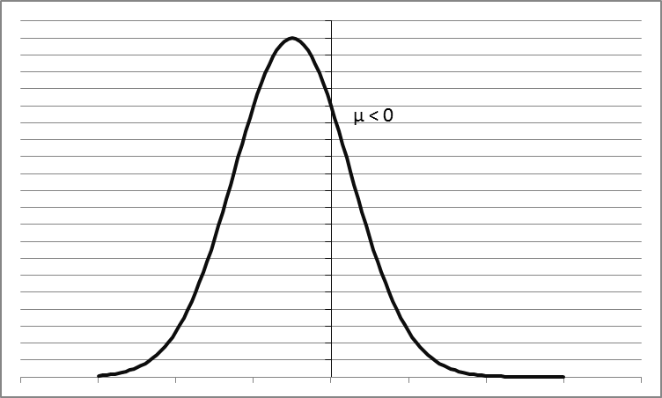
- relationships that relate y to the power of a base number, y = ax   
  
 e.g. y = 2x (already mentioned above)  
  
An important case of this is the exponential variety of relationships,   
involving the natural number ‘e’ introduced at the beginning of these notes.  
   
 e.g. the function y = eax   
  
shown below for various values of the parameter a (+/- 1 and +/-2).  
  
 

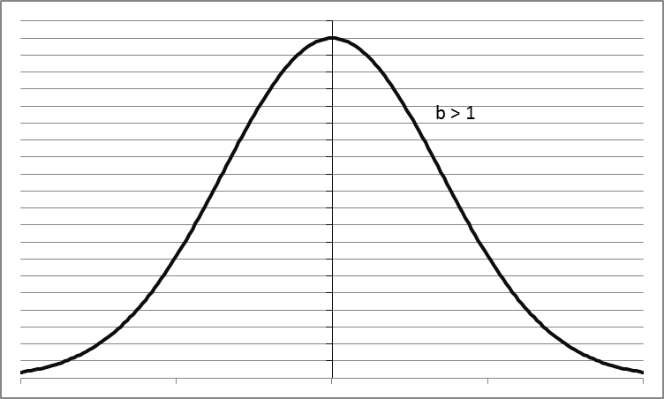
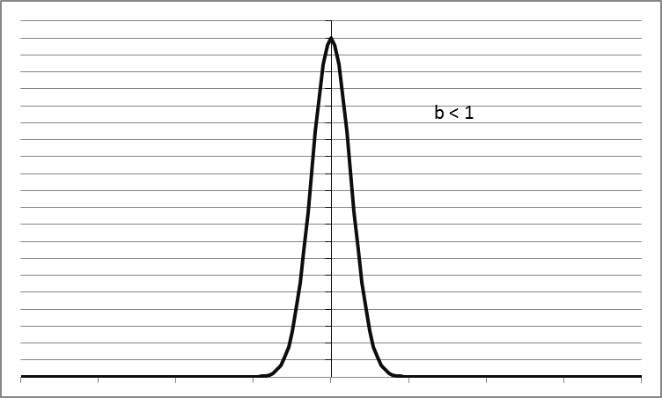
 

The function y = 1-e-ax (for positive values of a and of x)  
shows ‘growth’ starting from 0 when x = 0   
and gradually approaching the value 1 (or 100%) when x becomes large.   
It is often used to model growth phenomena or learning phenomena,  
which gradually approach a ceiling or maximum value in function of x  
and where x can stand for the passage of time, for expended effort or exercise.

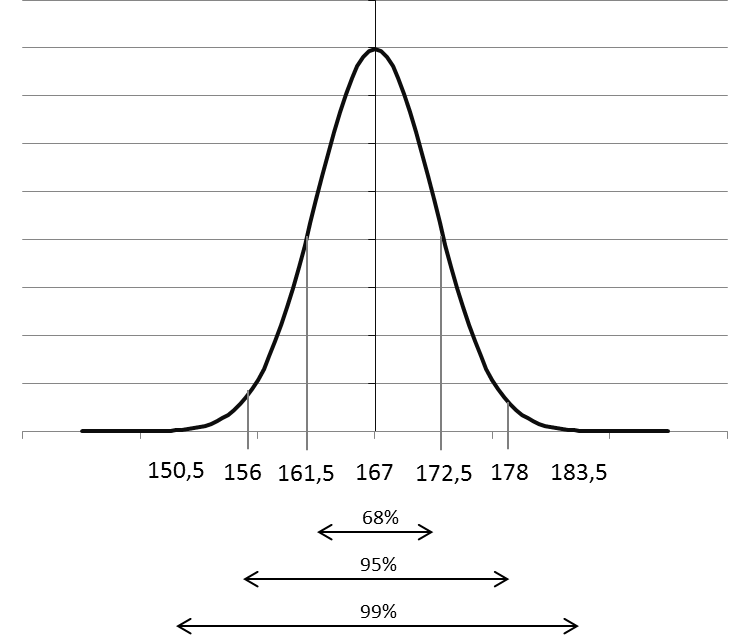
  
  
The reverse is the function y = e-ax,   
which starts at 1 for x = 0 and gradually approaches zero;   
it is often used to portray phenomena of forgetting or of decay.

  
- A growth phenomenon may start with an initial slow growth,   
followed by a phase of accelerating growth  
and next a phase of decelerating growth   
before ultimately reaching a ‘saturation level’.  
This relationship has an ‘S-shape’ and is called a **logistic or sigmoid** function.   
It can be represented mathematically as   
  
 y = xa/(b + xa)  
   
and is shown below for one specific values of the (positive) parameters a and b.  
A large parameter b has the effect of slowing the onset of growth  
(it shifts the curve to the right)  
a larger parameter a results in a sharper increase once growth has set in  
(it makes the function steeper).

  
  
This type of relationship is sometimes called a ‘diffusion’ curve;   
it is used to describe the spreading of a new phenomenon across a population,   
such as the spread of an infectious disease or the uptake of a new technology   
(e.g. the mobile phone, the tablet computer, Facebook membership, etc.).   
Such ‘diffusion’ usually takes a slow start, but then,   
as people ‘communicate’ the disease or the novelty to each other,   
catches up speed (‘snowballs’),   
and finally slows down and reaches a ceiling   
when all ‘susceptible’ members of the population have ‘contracted the disease’,   
or availed themselves of the innovation, or joined the social network, etc.  
  
If we have observations on the development of a diffusion phenomenon,  
we can estimate the parameters a and b by the least squares principle;   
*we demonstrate this with Excel in sheet 6.*  
 - a special and important exponential function   
is the **Gaussian or Normal** function,   
that of the bell-shaped curve mentioned above.   
Its formula is   
  
 y = a\*e-1/2{(x-µ)/b}^2  
  
with parameters a, µ and b;  
(note that this µ is not related to the µ used above; it is just a symbol)  
this curve reaches a maximum when x = µ;   
µ determines the location (the middle) of this symmetric curve  
the parameter b determines the tightness of the distribution   
the parameter ‘a’ is not relevant at this point.  
   
 

The bell-shaped curve represents the distribution   
of many phenomena in nature or in society,   
such as the distribution of heights in the adult population,   
the distribution of IQs in 10-year old children,   
the weight of ears of corn harvested from a field, etc.   
  
The parameter µ is the mean or average of the studied characteristic in the population   
(e.g. the average height, the average intelligence, etc.);  
the parameter b is the standard deviation of the distribution,  
(i.e. the typical difference of the characteristic in the population).   
  
An important property of the Normal curve   
is that fixed proportions of the population   
fall within specific intervals (expressed in standard deviations) around the mean,   
namely:  
- about 68 % of the observations lie within the range of +/- 1 standard deviation from the mean  
- about 95 % of the observations lie within the range of +/- 2 standard deviations from the mean  
- about 99 % of the observations lie within the range of +/- 3 standard deviations from the mean.

If, for example, the average height of the adult male population of Belgium is 167 cm   
and the standard deviation is 5.5 cm, then we know that  
- 65 % of the adult male Belgians have a height in the range of 167 +/- 5.5 cm  
- 95 % of the adult male Belgians have a height in the range of 167 +/- 2\*5.5 cm  
- 99 of the adult male Belgians have a height in the range of 167 +/- 3\*5.5 cm  
  
  
  
  
Even more detailed statements can be made   
about this distribution of heights in the population  
or about any other ‘normally distributed’ property of a population.   
  
A manufacturer of ready-to-wear clothing who would like to know   
what percentage of the population will need a particular size of clothes   
can rely on the Normal distribution to find this information;  
the Normal distribution could be used, for example, to determine  
how many male Belgians have a height between 160 and 162 cm.   
  
As stated above, the Normal distribution is one of the ‘laws of nature’   
that applies to many phenomena in nature or society.   
The Normal curve is one of the tools applied most often in statistics.

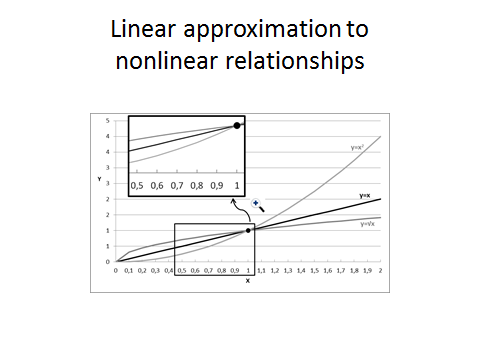
For a Normally distributed phenomenon   
(i.e. one that has this bell-shaped distribution)  
or for a phenomenon with any other known distribution,  
we can compute the probability of the occurrence of ‘events’ of interest to us.  
  
If, for example, we know that the IQ of a group is normally distributed   
with mean 125 and standard deviation 10,   
we may state that we expect the ‘event’ “a person will have an IQ above 155”   
(i.e. more than three standard deviations above the mean IQ)   
will have a probability of less than ½ % of occurring;  
stated differently, less than 5 people in 1000 will have that high an IQ.

As mentioned above, linear relationships are characterized by ‘**constant returns’.**Selling one more hat always yields an additional 50$ in revenue,  
irrespective of how many hats you are currently selling.  
Many, or even most, relationships are not linear however,  
and hence do not display ‘constant returns’.   
  
When I push in the gas pedal of my car, I drive faster,  
but the gasoline consumption also goes up.   
The relationship between speed and gas consumption is not linear however,   
i.e. it is not the case that for every km/h that I increase my speed,   
the gasoline consumption must increase by a fixed amount.   
Rather, as I drive faster, further increases in the speed of the car   
will consume progressively lager increments of gasoline.  
  
Another example of nonlinearity  
is that between size of cultivated land and agricultural output.   
A farmer first works the easiest and best land;   
if he wants to bring more land into cultivation,   
he will have to use progressively less fertile or less accessible pieces of land.   
He will also have to hire less productive hands to help out.  
This means that, as he brings more land into production,  
as he employs more workers,   
the increment in production he obtains from using more land and more labour  
(while still positive) becomes progressively smaller.   
The latter is called a phenomenon of   
‘**decreasing incremental returns**’ or ‘**decreasing marginal returns**’.

Given the amount land already under cultivation,   
a further increment of inputs yields progressively smaller increments of outputs.

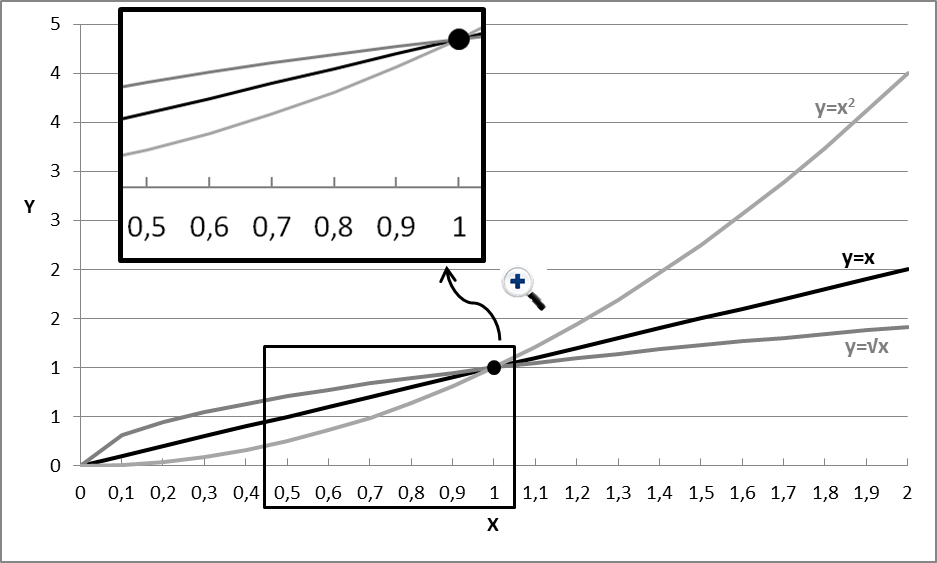
Sometimes we witness **increasing marginal returns**   
instead of decreasing marginal returns.   
Consider a social network (such as Linked In);   
since the value of such a network is to be a ‘network’,  
its value depends on its number of members:   
if the network has only one member, it is of no value;  
if the network has two members, A and B,   
then only one link can be established (between A and B);  
that is not much of a value…   
with three members, A, B and C, we have 3 links: AB, AC and BC.   
With four members there are 6 links (AB, AC, AD, BC, BD, CD);   
with 5 members 10 links, with 6 members 15; etc.   
This is a bit like with logarithms (see above):  
as the number of members increases from 1 to 5, etc.  
the number of links ‘explodes’ from 0 to 15, etc.  
A network with n members in fact allows for n\*(n-1)/2 links.   
So, with 10 members, a network allows for 10\*9/2 = 45 links,   
and with 20 members for 20\*19/2 = 190 links.   
  
The number of links increases proportionally with n² (i.e. in a quadratic way),   
rather than linearly (proportionally with n).   
This is an example of increasing marginal returns:   
as the network grows, every new members adds a steadily increasing number of links.  
The returns to additional members keep on increasing ever faster.   
If you have the choice between two networks,   
one with fewer and one with more members,   
other things being equal, you should opt for the larger network,   
because it will give you more communication links;   
but if you join the larger network, then that network becomes even larger,   
and hence even more attractive for new members to join…  
   
In the end, only one network can remain;   
economists call this a situation of ‘**winner take all competition’**.   
In a situation of increasing returns to effort,  
only one competitor finally survives and takes the whole market.   
That, by the way, is why economists tend to dislike situations   
characterized by increasing returns to scale or increasing returns to effort:   
in the end this must lead to a monopoly situation,   
where there is no competition anymore,   
a situation where one supplier calls all the shots.

The graph below shows an example   
of three types relationships between inputs and outputs:   
(1) constant marginal returns (e.g. y = x),   
(2) decreasing marginal returns (e.g. y = x1/2  and   
(3) increasing marginal returns (e.g. y = x²).   
In the latter two cases, we can clearly see the nonlinearity of the relationship.



**4.6 Linear approximation of nonlinear relationships.**

Nonlinear relationships are, by definition, not linear.  
  
But when we focus on a small interval of the independent variable (x),   
nonlinear relationships often appear as almost linear.  
In such a small range of x, a linear relationship y = a+bx   
may quite satisfactorily approximate the nonlinear relationship.   
  
In the picture below we see   
that the relationships y = x1/2  and y = x2 are nonlinear in the interval 0<x<2.  
But in the window, we see that these relationships are almost linear  
in the interval 0,5<x<1.



In a sufficiently small interval, any relationship appears almost linear,   
and can be dealt with as such (i.e. as linear) for most practical purposes.

Here is an example of what we mean by linear approximation in real life.   
We know that earth is a sphere: the surface of the earth is not flat.   
But when we lay railroad track, the sections are so small,   
compared to the surface of the earth,   
that we can use straight (linear) rails;   
the track is laid in ‘successive linear approximations’.   
  
On the other hand, when an airline pilot charts a long-distance flight,   
he will take into account the curvature of the globe.   
The distance he flies is not negligible relative to the surface of the globe;  
in this case the nonlinearity really matters.   
  
The conclusion is that linear approximations to nonlinear relationships   
may suffice in many cases, but not in all,   
and that linear approximations may suffice   
as a first approach to studying many relationships.

**Linearizing nonlinear relationships.**

In some cases , when y is a nonlinear function of x,  
it is possible to transform y in such a way   
that the transformed values are a linear function of x.  
  
In the Genesis example above:  
Development (day in genesis) Number of years in the past Ratio to next number

1. Big bang, creation of the universe 15.000.000.000 1,94  
2. Stars, our sun and planets form 7.750.000.000 2,06  
3.Earth cools, water and dry land appear 3.750.000.000 2,14  
4. Earth atmosphere forms 1.750.000.000 2,5  
5. First animals appear 700.000.000 2  
6. Mammals, hominids and humans appear 350.000.000

the number of years between the phenomena (y)   
increases with a ratio 2 for each increase by one day in the number of biblical days (x);   
real time y clearly is a nonlinear function of biblical days x.  
We have shown above that if a phenomenon grows at an increasing rate  
then the logarithm of its value increases only linearly.  
In the case of Genesis, the dates corresponding to the days of creation  
seem to evolve as 15 000 000 000\*2X , with x = 0, -1, -2, -3, -4, -5, -6   
With y the real passage of time, and x the number of the day in Genesis,  
the logarithm in base 2 of real time is a linear function of x.  
We write this as log2(real time) = a + bx, a linear function of x.  
  
More generally, if a phenomenon y grows   
exponentially as a function of x  
for example y = ea+bx ,  
then y is clearly a nonlinear function of x,  
but since ln(y) = a+bx,  
the logarithmic transform of y becomes a linear function of x.

The operation of taking the logarithm   
transforms the nonlinear relationship between y and x   
into a linear one between ln(y) and x.  
Whereas the phenomenon evolves as a nonlinear function of x,   
its logarithm evolves as a linear function of x.  
  
If you are a non-mathematician, you should be thoroughly confused by this;   
but the message is the following:   
if you study a phenomenon y that is clearly a nonlinear function of x;  
there might be transformations of y that are a linear function of x.  
  
Linearizing nonlinear relationships offers several advantages:   
linear relationships are easier to understand and represent;   
statistical theory and tools for dealing with linear relationships   
are better developed and efficient than those for nonlinear relationships.  
  
That is why we sometimes prefer to transform nonlinear relationships  
into linear ones.

1. **Some useful geometry and trigonometry.**

Angles, squares, rectangles and triangles hold few secrets for us.

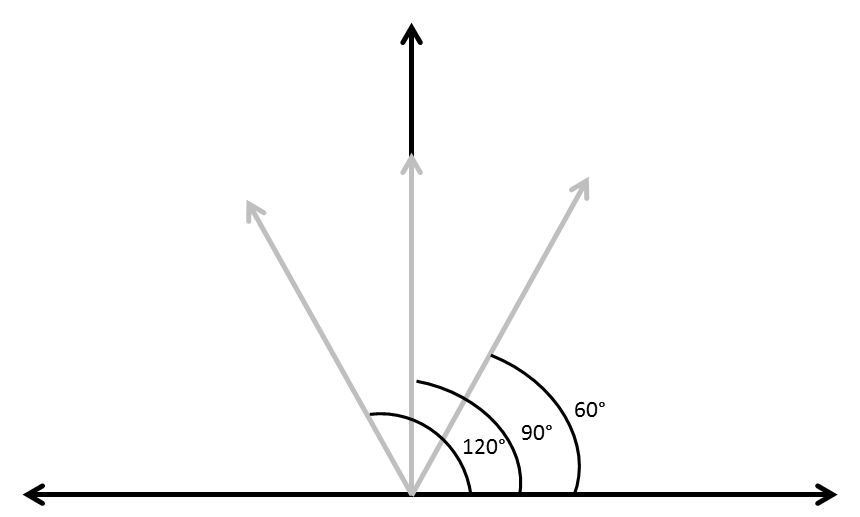
**Angles**

Angles are measured in degrees (using a number system with base 12…)   
a full circle measures 360°  
(360° is 4\*90° or 2\*180°, numbers which will often appear again below).  
  
Remember that the circumference of a circle   
is given by the formula 2\****pi\****r or (6,28…)\*r,  
where r is the radius of the circle.  
***pi*** or 3,14…. is not a ‘nice’ number to have to deal with,  
but as we explained, it happens to be fundamental to our universe.  
If you do not like it, don’t blame mathematics, blame God!

**The next thing is unimportant, but useful to know.**

The circumference of (6,28)\*r of a circle corresponds a full circle of 360°,  
hence (6,28…)\*r = 360°  
Using the basic rules of algebra, we can then say that   
r = 360°/(6,28….) = 57,296°  
A distance of r along the circumference of a circle  
corresponds an angle of 57,296°.  
That is also called a ‘**radian**’  
So, you can express angles in degrees or in radian,   
it does not matter,   
just like you can express temperature in °C or in °F.  
Mathematicians prefer to express angles in radians,  
so that they can also work in a decimal system,  
like they do with meters, liters, kg, etc.  
**You only need to know this  
because we will use angles to explain some things below  
and Excel works with radians rather than with degrees.  
So, our Excel examples will use radians instead of degrees**

A ‘**right angle’** measures 90°  
(which is also 90/57,296 radian or 1,57 radians).   
It is also called ‘**orthogonal**’   
(orthos in Greek means ‘right’, or ‘at right angle’; gonios is angle).  
  
A ‘**sharp angle**’ is one between 0° and 90°  
a ‘**blunt angle**’ is one between 90° and 180°

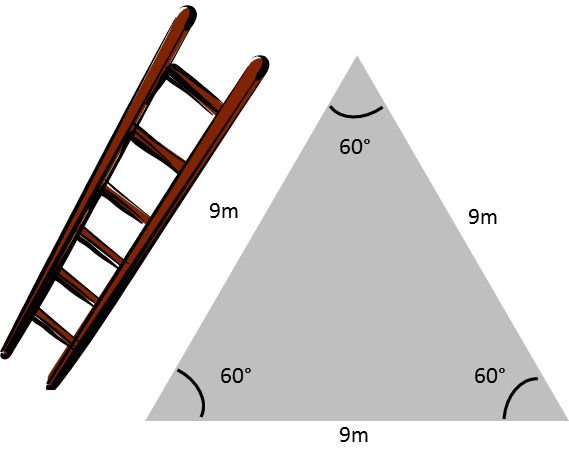


**Squares and rectangles**

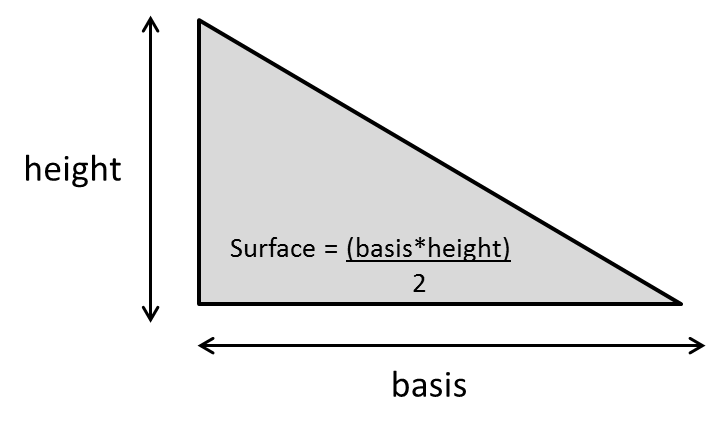
A square has four equal sides.  
A square with sides of length ‘a’ has a circumference of 4a and a surface of a\*a or a².  
Hence , a square with a side of 1 meter has a surface of 1 square meter (1m²)   
and a circumference of 4 meter;  
a square with a side of 2 meters has a surface of 4 square meters (2m²);   
if the side of a square is half a meter (0.5 meter or ½ meter),   
then its surface is one quarter of a square meter: 0.5² = (1/2)² = 0.25m² or ¼m².  
  
A rectangle with sides of length a and b has a circumference of 2a + 2b  
and a surface of a\*b.

**Triangles**

A triangle has three sides and three angles;   
the three angles **always add up to 180°.**  
If you know the size (in degrees) of two angles,   
then the third angle is 180° minus the other angles.  
If, for example, one angle measures 90° and the other 60°,  
then the third measures 180° minus (90° + 60°) or 30°.  
  
If you know two of the three angles of a triangle  
(then you know the third: as 180° minus the other angles),  
you can draw the shape of that triangle (how it looks),  
but not its size (how large or small it is).  
But if you also know the length of one of its sides,  
then you can draw it perfectly, both shape and size.

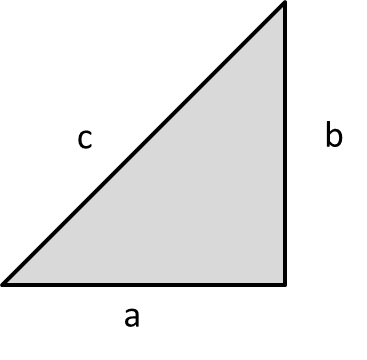
This shows that you do not need to know   
the size of all the angles and of all the sides  
to know a triangle exactly,  
e.g. if you know two sides and one angle  
or two angles and one side  
that is enough to draw the triangle exactly.   
  
That knowledge can be useful.   
Let us say that I want to replace the shingles of my roof,  
and to do that I need to buy a ladder  
which I can lay against one of the sides of the roof.  
My roof is symmetric, has a pointed angle of 60° and a base of 9 meters.  
  
How long a ladder must I buy?   
Since the roof is symmetric,   
its two base angles must equal (180°-60°)/2 = 60°.  
Since all three angles are equal, all its sides must be equal too  
and measure 9 meters.  
Conclusion:   
I must buy a ladder of 9 meter long…  
  


The circumference of a triangle is the sum of the lengths of its three sides.  
The surface of a triangle is given by   
one half the product of its basis and its height: (b\*h)/2.



If, for example, the Gardening Services of the city wants to compute   
how many plants they will need for a triangular park in the town,   
with the base of the triangle of 20 meters and the ‘height’ 15 meters,  
and 20 plants are needed per square meter,  
then the surface of the park can be computed as (20m\*15m)/2 = 150m²  
and they will need 150\*20 = 3.000 plants.

**Rectangular triangles**

A triangle with one angle of 90° is called **rectangular.**  
A rectangular triangle has one longer side, opposite the right angle,  
(let us write its length as ‘c’)  
and two shorter sides, its basis and its height  
(with lengths ‘a’ and ‘b’)  
We will usually represent the sides as in the picture below,  
for reasons that will become clearer as we progress.  
  
**Here is a very, very important and useful finding concerning rectangular triangles:   
 for any rectangular triangle, with long side of length c   
 and short sides of lengths a and b,   
  
 c² = a² + b²**  
The proof of this was given by Pythagoras;   
it is therefore called **Pythagoras’ rule.**

If you are interested in the proof,  
you can verify this with a rectangular triangle with sides a and b of equal length.  
If you are not interested in the proof, just move on!  
  
Proof:  
such a triangle is the half of a square with side a;   
its surface must therefore equal a²/2 (half the surface of the square, a²).   
Now, c is the length of long side of the triangle,  
and the height of that triangle, using c as its basis must be c/2.  
Hence the surface of the triangle must be c²/4.  
Hence, we find that a²/2 must equal c²/4,  
and hence c² = 2a² = a² + a², which is what we wanted to demonstrate.  
  
**Pythagoras’ rule has numerous practical applications,   
not only in the physical world (e.g. to build pyramids)  
but also theoretically or conceptually.  
Especially because a² + b² is a sum of squares,  
and we know that sums of squares are important to us.   
That is why we need to discuss this here**.

If, for example, city A lies 20km East and 30 km North of you,  
then, what is your distance to A (‘as the bird flies’)?  
If we make a picture of this, then you and A are the endpoints   
of the long side (c) of a rectangular triangle,   
with short sides of 20 km (a) and 30 km (b).  
With the rule c² = a² + b² = 20² + 30² = 1300,   
we find that c = 13001/2 = 36,055 km.  
  
Likewise, if there are two cities, A and B  
and B lies 50 km East and 80km North   
and B lies 30 km East and 120 km North of where I live,  
then what is the distance between A and B?  
We can again draw a rectangular triangle, with long side c  
and with short sides (50-30) and (120-80).  
The answer then is c² = (50-30)² + (80-120)² = 20² + 40² = 2000  
and hence c = SQRT (2000) = 44,72 km.

Likewise, if there is a point that lies 1,5 meter in front of me,   
and 3,2 meter to my right and 2,4 meter above me,  
then the distance between me and that point will be  
SQRT(1,5² + 3,2² + 2,4²) = 4,272 meter.  
This is a simple extension of Pythagoras’ rule   
to a third dimension.

You can see, then, that Pythagoras’ rule can be used  
to compute real physical distances.

But there is more to it than that:  
in the formula SQRT(1,5² + 3,2² + 2,4²)   
we recognize something:  
it is the square root out of a sum of squares!  
  
Since the standard deviation of a series of numbers  
also is the square root out of a sum of squares,  
we will use that correspondence below  
to draw analogies between series (vectors) of numbers,  
their standard deviation, variance and other characteristics  
on the one hand  
and distances, lengths, and other concepts  
in graphical representations of data or measurements.   
  
More of that later;   
for now, we just want to point out the many uses of Pythagoras’ rule

Another use of the rule is to decompose and represent  
the fluctuations (i.e. its variance) in a variable (c²)  
into that part of the fluctuations   
that is ‘shared with’, or can be explained by one or several other variables (a²)  
and the part that cannot (b²).  
  
The weight of a child, for example, tends to (co)vary together with its height;  
stated differently, differences (or variance) in weight between children (c²)  
can in part (but not completely)   
be explained by differences (variance) in height between the children.  
The part that is explained is represented as a²  
and the part that is not explained as b²  
and c² = a² + b².  
We can then also say that   
“(differences in) height can explain (a²/c²)% of (the differences in) weight”.  
  
 c² (differences in weight) =   
 a² (part of differences in weight explained by height differences)   
 + b² (part of differences in weight not explained by height differences)

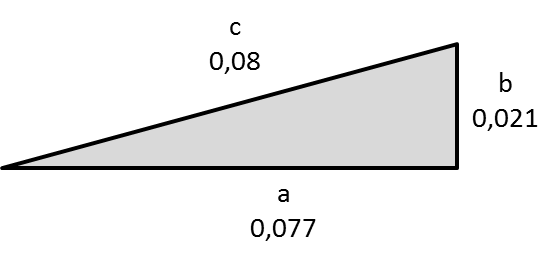
This leads us to a further application of Pythagoras’ rule:  
the concept of the **reliability** of a measure or measurement instrument,  
for example, the reliability of my bathroom weighing scale.  
  
I watch my weight; every morning, to see if my weight is under control.  
Now, no bathroom scale is absolutely perfect:   
if I measure my weight five successive times one particular morning,   
I might read the following numbers (in kg) on the scale

64.75 64.71 64.76 64.72 64.71

The average of my weight measurements that morning is   
 (64.75 + 64.71 + 64.76 + 64.72 + 64.74)/5 = 64.73 kg  
The variance of those numbers is small: 0,00044

That my bathroom scale does not always give the same number   
does not bother me too much:   
the fluctuations that I observe are minor;  
I understand that there are multiple influences, small and irrelevant (random)  
that affect each specific measurement or observation:  
how I step on the scale, where I put my feet on the scale,   
the temperature of the room, the humidity in the room, etc.)   
Small fluctuations are possible:  
the variance of these five weightings is only 0,00044,  
their standard deviation or typical difference is only 0,021 kg.  
These are ‘random’ fluctuations I can live with;   
there is no need to throw out my bathroom scale because of this.  
  
Now, if on four successive weeks   
my weight is 64.70 kg 64.85 kg 64. 81 kg 64. 92 kg  
then it seems like my weight is increasing!   
Should I panic, or might these differences   
be due to random measurement fluctuations as just discussed?  
  
The variance of these four weightings is 0,0064.  
That variance (call it c²) reflects two sources of fluctuations:  
real fluctuations in my weight from one day to the next (if there are any; call them a²)  
and random fluctuations that affect every measurement (call them b²).  
If we apply the rule c² = a² + b²,   
or total differences : real differences + random differences  
we can compute that   
a² (real flucutations in weight) = c² - b² = 0,0064 – 0,00044 = 0,00596.  
Hence the reliability of my scale would be 0.00596/0,0064 = 0,93 or 93%  
My scale seems to be quite reliable,   
it is therefore probable that I have really picked up weight.

A pictorial representation of this problem is given in the graph.

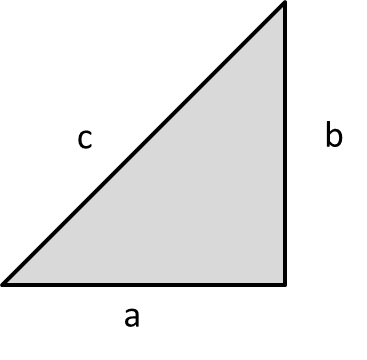


The ratio of non-random fluctuations tot total fluctuations in a measurement  
is called ‘**the reliability**’ (of the measure or of the measurement instrument),  
i.e. the percentage of the fluctuations in the measurement  
that is not due to random fluctuations,  
i.e. a²/c² = 0.00596/0.00640 = 0.93 or 93 %.  
I can say that my bathroom scale is ‘93% reliable’; I can ‘trust it for 93%’.

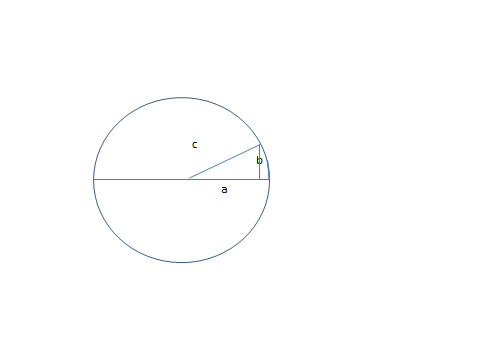
Just as a reminder: all of this is an application of Pythagoras’ rule.  
So, thank you Pythagoras!  
  
Pythagoras’ rule will be used often below  
to discuss data analysis methods  
and to explain them in a graphical way.

**Trigonometry**

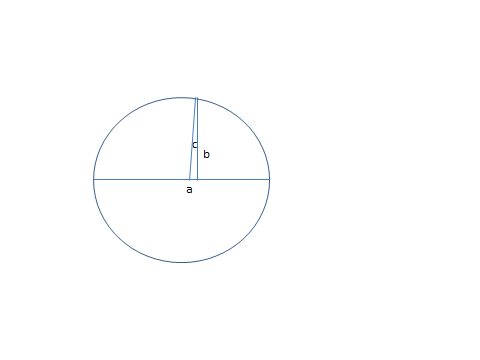
Trigonometry – the study of triangles-  
will sound like an esoteric subject to many of you.  
We will not need it for its own sake,   
but in order explain data analysis in a graphical way.

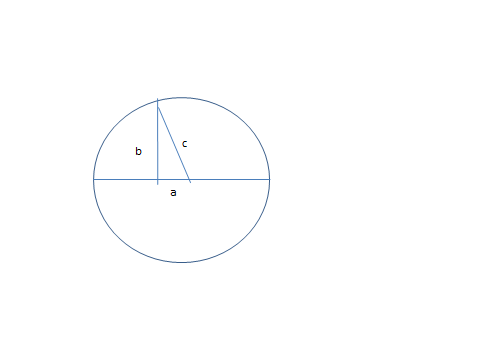
Let us revert to the example used above  
where we discussed to what extent the (differences in )weight of children  
could be explained by means of (differences in) their height.  
  
This led to the use of a rectangular triangle,  
with c² representing variance (differences) in the weights,  
a² the share of those differences   
that can be predicted (or ‘explained’) by differences in height  
and b² the share of the variance c² that remains unexplained.  
  
If the angle between a and c (in our example weight and height)  
would be 45°, i.e. half-way between 0° and 90°,  
(or halfway between totally flat and completely steep),  
the a and b would be equal, c² would equal 2a² and a²/c² would be 0,50.  
That would then represent a situation where  
‘half of the differences in weight between people  
can be explained by the differences in their height’.  
  
We can also compute a/c = SQRT(0,50) = 0,71.  
The latter number will be very useful below.

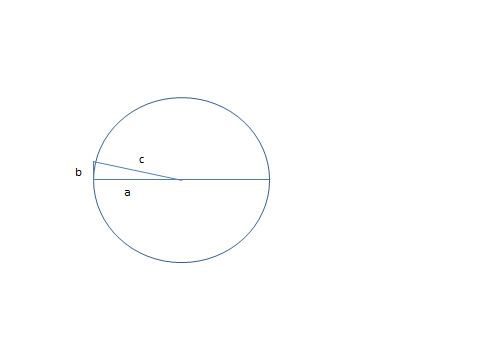
If the angle is sharper, e.g. 30°,   
a will be larger relative to c,  
and we can compute (how that is done is not relevant here)  
that a²/c² = 75 % and a/c = 0,87.  
This would mean that both concepts or measures,  
in our example height and weight  
would be strongly related,   
that height could predict weight with 75% accuracy,  
that height and weight seem to have a lot in common.



When the angle approaches 90°,   
let us say 85°, then side a becomes small relative to side c  
and one can compute that a²/c² is only 0,01

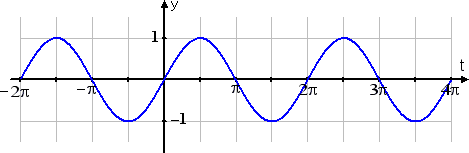


It is easy to see that when the angle is 90°  
a = 0 and hence a²/c² and a/c = 0.  
This would mean that the two concepts or measurements, a and c  
share no information, have nothing in common.  
They would then be called ‘**orthogonal**’ or ‘**independent**’.  
This we would not expect between weight and height, of course,  
but you might expect that to be the case between, for example,  
the intelligence (I.Q.) of people and their height:  
how tall people are and how they score on an intelligence test  
is unrelated   
(at least, that is what I think; to be really sure, we should test that with real data).  
  
We can go a bit further, and look at the picture   
for an angle larger than 90°, e.g. of 120°.  
  


We see that a lies in a direction opposite to that of a in previous pictures.  
a/c now is a negative number, actually -0,5.  
That would express that a and c have something in common,   
but in an opposite sense:  
if c would be weight and a would be height,  
this would mean that those who are taller actually weigh less.  
Of course, that is not something we would expect   
with weight and height data.  
But you might encounter that if the two variables  
were weight on the one hand   
and number of hours/week devoted to sports on the other hand.  
Even though a/c is negative (-0,5)  
a²/c² is a positive number, expressing that in this case  
(-0,5)² or 25% of the differences in weight   
would be explained by the opposite of height   
(by ‘smallness’ rather than ‘tallness’).  
  
Finally, if the angle between a and c would be (close) to 180°,  
then side a would become large (again) relative to c;   
a/c would be (close to) -1.0  
and a²/c² close to 100%.  
With weight and height, this would mean that you can (almost) perfectly predict weight   
from the reverse or opposite of height (i.e. the ‘smallness’).  
Again, that is not likely to happen in reality,  
but it could be a result that you would obtain, for example,   
between the weight of a person and his/her intensity of dieting.

The ratio a/c is called the **‘cosine’** of the angle between a and c,  
it actually expresses the direction (i.e. positive or negative)  
and strength (between 0,00 and 1,00) of the relationship  
between two concepts or measurements,  
such as e.g. weight and height.  
Another name for it is **‘correlation’**  
Correlation is absolutely a key concept in data analysis;  
we will explain below how to compute it.  
  
As you see, Pythagoras, with his rule  
has allowed us to think and represent in pictures  
relationships between concepts (variables, measurements),  
to picture the direction and intensity of their relationship.  
Thank you, Pythagoras!

Just before we move on, let us ad a little ‘bonus’ for you.  
In the circle we have used to represent c, a and b,  
you could continue the exercise for angles beyond 180°,  
for example of 200°, of 290° or 350°,  
until, ultimately we reach 360° and are ‘full circle’.  
  
If you would compute the cosine corresponding to each such angle,  
you would see that the cosine evolves from a value of 1.00 to a value of 0.00,  
next to evolve to -1.00, then back to 0.00 and finally 1.00 again.  
  
You obtain a relationship   
between the angle x and its cosine, y = cos(x)  
like in the picture  
(the picture is not really exact, but do not worry…).  
Now we see that Pythagoras’ rule even allows us  
to model and analyze phenomena that take on the shape of a wave,  
e.g. waves in the water, sound waves, radio waves,   
brain waves, waves of economic activity or whatever waves you may encounter.  
If you are studying a phenomenon that behaves like a wave,  
there are good mathematical tools to represent and analyze them!



1. **Data tables and matrices.**

We are now ready to discuss one of the most important reasons  
for this course of practical mathematics:   
the treatment of data tables, the processing of data.  
  
In many cases, research will involve measuring a number of characteristics,  
e.g. height, weight, intensity of sports practice (1= low, 5 = high )  
of a number of entities.  
This results in a data table such as in Excel sheet 7,  
with 3 v ariables (columns) and 8 entities (children - rows)

**6.1 Vectors of observations or measurements**

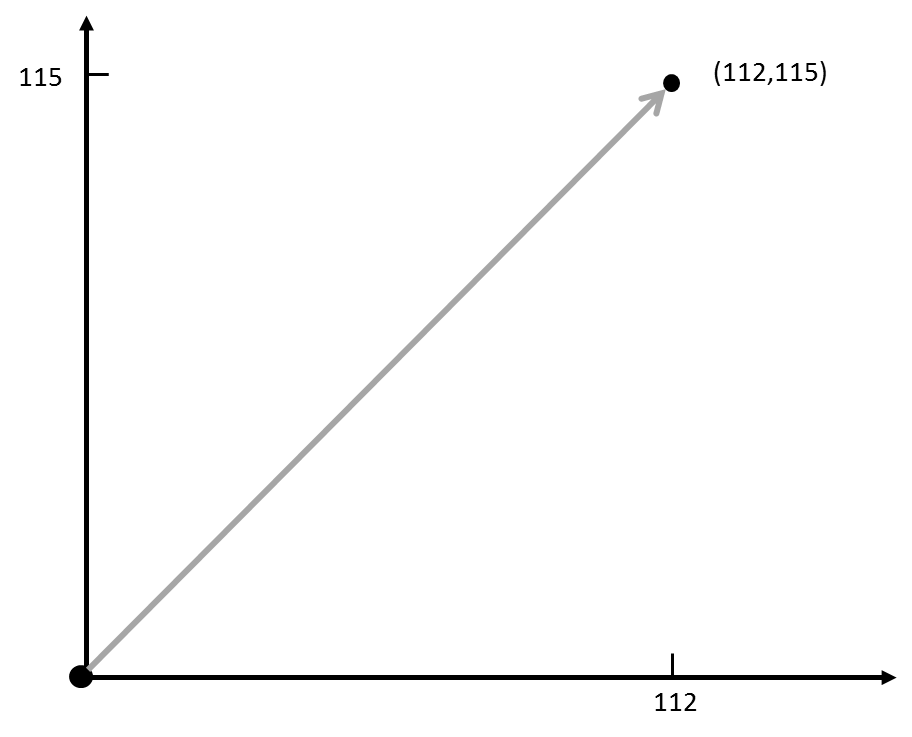
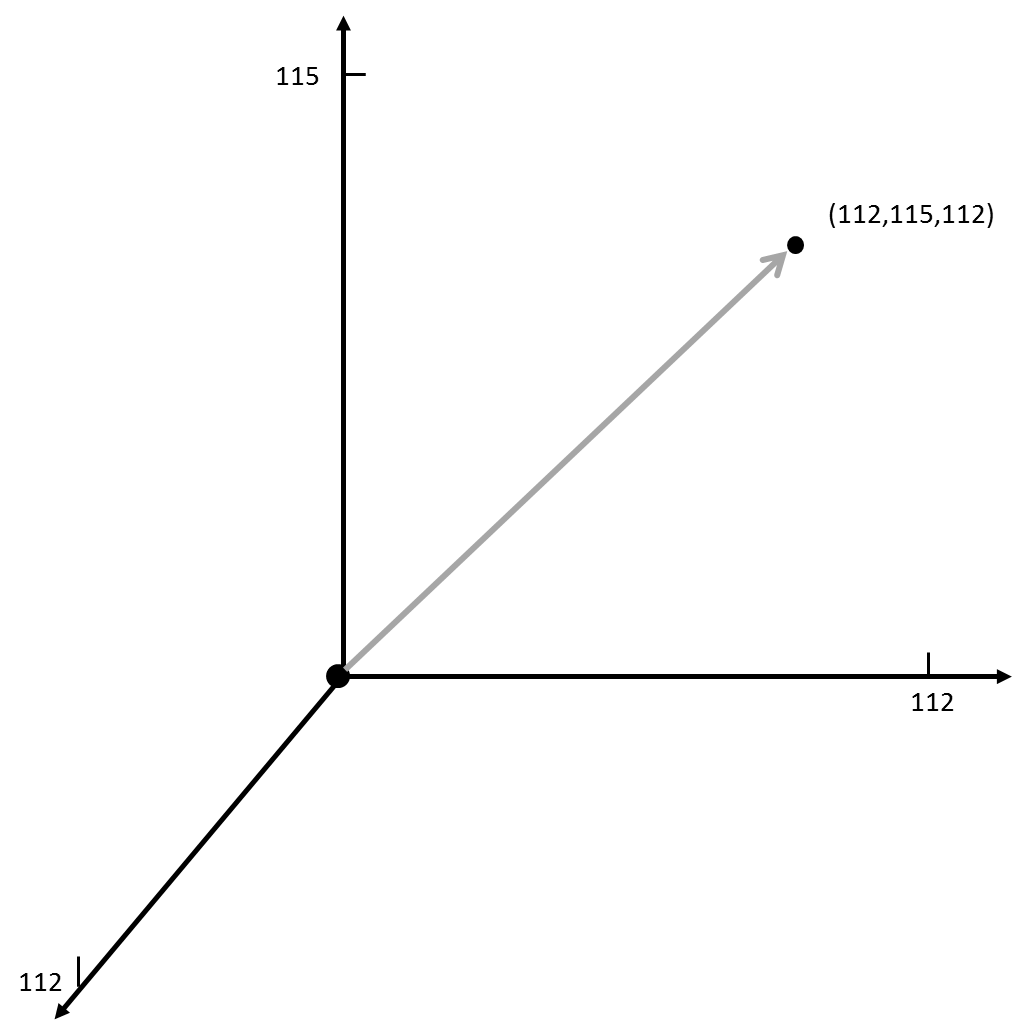
The heights are H’ (112, 115, 112, 106, 120, 105, 111, 109).  
  
A series of numbers or observations such as this is also called a ‘vector’,  
typically showing the numbers stacked from top to bottom (H),  
or the transpose of the vector, with the numbers from left to right (H’).  
  
We say that H has dimension (8\*1), i.e. eight rows of one number each,  
and H’ then has dimension (1\*8), i.e. one row of eight successive numbers.  
  
H’ and H can then be seen to be ‘commensurate’,  
i.e. H’\*H can be ‘multiplied out’ as a sum of squares   
of (1\*8)(8\*1) = (1\*1) = 1 number,  
if the two inner numbers are the same (8 and 8 in this example).

112, 115, 112, 106, 120, 105, 111, 109 112  
 115  
 112 = 112² + 115² + 112² + 106²   
 106 + 120² + 105² + 111² + 109² = 99176  
 120  
 105  
 111  
 109

Likewise, with two different data series or vectors,   
one for the heights (H) and one for the weights (W),  
we can compute the sum of the ‘cross-products’ of Height and Weight,  
written as H’W   
*as shown in column O of Excel sheet 7.*

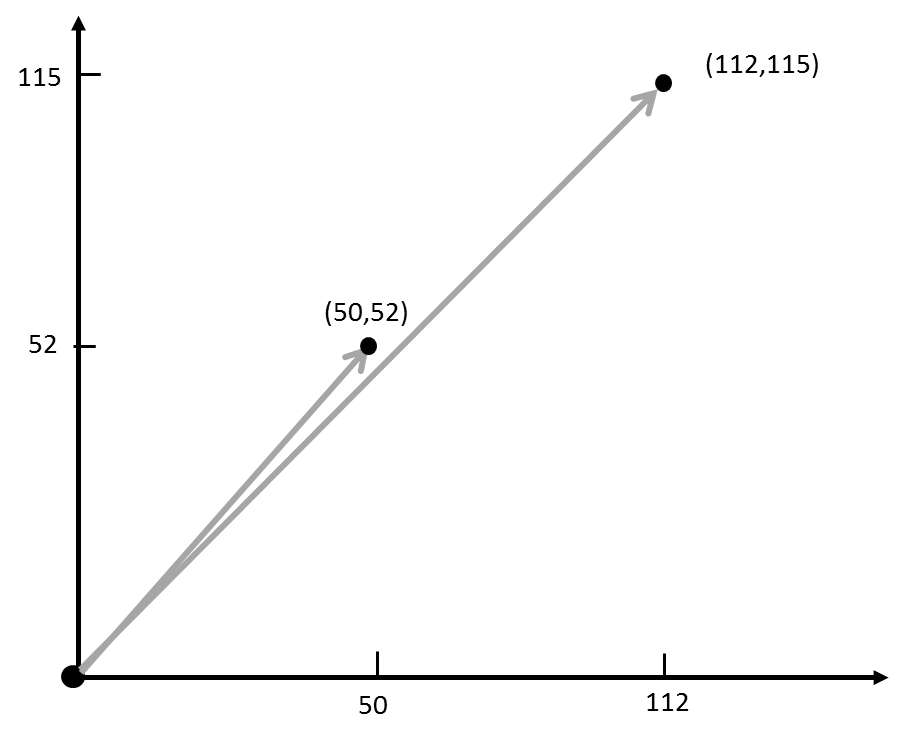
The ‘length’ of a vector   
is obtained by taking the sum of the squares of its elements  
and then taking the square root of that sum   
(the square root of a sum of squares again!  
  
That, in fact, is the formula used above   
to compute ‘the distance between yourself’   
and city A using Pythagoras’ rule.  
If city A lies 112 km to the East and 115 km to the North  
(we choose the same numbers as the first two height numbers for illustration  
and to show something later on).  
then the distance (‘as the crow flies’) from you to A,  
by Pythagoras’ rule, is   
SQRT{(112-0)² + (115-0)²},   
i.e. the length of the vector from you, the origin (0,0)   
to its endpoint (112, 115)  
  
In the case of the vector of heights, therefore, the length is  
SQRT (112² + 115² + 112² + 106² + 120² + 105² + 111² + 109²) = 314.92  
  
Note that if you divide all the elements of the vector by its length, 314.92,   
the resulting vector will, necessarily, have length 1.0

If you have three or more numbers in the vector,   
the formula remains the same.   
  
Vector actually means something line ‘direction’ or ‘arrow’.  
It is useful to think of a data series or data vector  
as an arrow of a certain length (you know how to compute that now)  
and pointing in a given direction (more about that immediately).  
  
If you want to ‘see’ this in the case of the vector of weights,   
picture the following:  
take a ruler, draw a (first) line   
and measure off the first weight, 112, along that line.  
Now draw a second line at right angles with the first   
(both lines start from the same point or ‘origin’)   
and measure off the second weight, 115 on it.   
The line (or arrow) which connects the origin   
with the point with coordinates (112, 115)   
is the vector corresponding to the first two observations.

  
  
Now (imagine that you) draw a third line at right angles to the first two  
(it will stick right out of the page on which you drew the first two axes)  
and measure off the third weight, 112, on that axis.   
  
You can again draw the vector linking the origin and the points (112, 115, 112),  
the vector made up of the first three observations.  
Its length, i.e. the distance from the origin or the point (0,0,0)   
is SQRT {(112 – 0)² + (115 – 0)² + (112-0)²} or SQRT {112² + 115² + 112²}   
  


Note that if you were a ‘being’ that lives only in two dimensions,  
one that knows only a flat universe, with no ‘up’ or ‘down’,  
you might consider this thing with the third dimension weird, hard to picture.   
  
For us, humans, reasoning in three dimensions is not a problem.   
Moving to four dimensions is a bit more of a problem to us, though:   
but just imagine (not in pictures, but ‘logically’) that you draw a fourth line,   
‘at right angles’ with the first three,   
and you measure off the fourth number, 106, along that line.   
  
You can now imagine drawing the arrow   
from the origin to the point with coordinates (112, 115, 112, 106).  
You cannot picture this with your eye or with your visual imagination;   
but you understand that it is just the same as in three dimensions,   
but now with one dimension added.   
It just requires some mathematical imagination or hallucination   
(or the consumption of drugs…) to ‘see’ that.  
But mathematically, with our mind,  
it is just the same, but in one more dimension.

And so we can continue,   
until we represent our series of eight numbers   
as a vector or arrow from the origin   
to the point with coordinates (112, 115, 112, 106, 120, 105, 111, 109)   
in a ‘space of 8 dimensions’   
and with length SQRT {112² + 115² + 112² + 106² + 120² + 105² + 111² + 109²}

We have represented the numbers for height ‘graphically’  
as an arrow, a direction,   
pointing out from the origin in a space of 8 dimensions  
and with length as given above.  
  
What if we had another vector,   
with numbers for another property of the 8 children?   
Let us say that we have measured also their weight,   
with the following (transpose) data vector for the weights  
 W’ = 50 51 46 49 54 51 59 55  
  
Now let us enter the 8 numbers for weight   
in the same 8-dimensional picture used to represent the height numbers.   
  
We now see two arrows:   
one representing height and the other weight.  
  
We show how this would look with only the first two coordinates  
for height and weight.  
  
   
As both vectors or arrows start from the origin (0,0),   
these two arrows make an angle with one another.   
  
We repeat a conclusion from the previous section:  
if two variables, e.g. height and weight, are associated positively   
(i.e. taller people tend to weigh more,  
or conversely, the more you weigh, the taller you tend to be),   
then the angle between these two vectors will be rather sharp  
(between 0° and 90°).   
If height and weight would be unrelated, the angle would be around 90°   
(i.e. a right angle, both vectors would be ‘orthogonal’);   
And finally, if height and weight would be associated negatively,   
i.e. would tend to move in opposite direction,   
(i.e. taller people weigh less, smaller people more – obviously not what one would expect)   
then both vectors would make a dull angle (between 90° and 180°).   
  
Of course, we expect height and weight in children   
to bear some positive relationship to one another   
(hence their relationship to be rather close, or their joint angle rather sharp);   
but we do not expect height and weight to be perfectly related   
(which would be represented by a perfect overlap between both vectors);   
indeed, weight is caused not only by height, but also by other things,   
such as tendency to overweight, extent of exercise,   
and many other, often rather random influences,  
so that the overlap between both could never be perfect,  
or, using the symbols introduced above, a²/c² < 1.00.

We have now shown again, but with a different approach,  
that two (or even more) vectors of numbers  
can be represented visually as arrows   
pointing into space from a common origin,  
each with its own length  
and with the angle between them  
expressing the direction (i.e. positive or negative)   
and the degree (i.e. weak or strong)  
of association between the variables they represent.

* 1. **Vectors of centered data.**

Instead of working with the original data for height and weight,  
we could work with the centered data.  
Nothing much will change with the children and their height or weight:  
they will still measure and weigh the same  
and the association between height and weight will be unchanged.  
  
But the picture will differ a bit:  
the endpoints of the arrows can now be positive or negative  
(as is the case with centered values)  
and the length of the arrows will now be   
the square root out of the sum of squared centered values.  
  
The square root out of the sum of squared centered values,  
the length of the vector of centered values,  
is the same as the formula of the standard deviation of the (original) numbers  
(with one exception: for the standard deviation we use   
the average of the summed squares instead of only the summed squares;   
that is purely to place vectors with different numbers of observations   
on equal footing   
(e.g. to compare the standard deviation of a series of 8 heights  
with that of a series of e.g. 20 heights).   
  
Centered values can therefore, again, be represented as arrows  
(in 8-dimensional space in our example),  
with length corresponding to the standard deviation of the numbers.

If the numbers are all the same, then the centered values are all the same,  
then the arrow representing them would collapse in the origin  
(they would all be zero)  
and the standard deviation (and also its square, the variance)  
would also be zero (i.e. a sum of squared zero’s);   
the vector would have no length.  
  
The length of the vector, therefore,   
expresses the extent to which the numbers that it contains differ.  
  
We also say that a vector with standard deviation or variance zero  
‘conveys no information’.  
It is a little bit like with music:  
if the notes of a piece of music would consist   
of only one tone, we might say that it ‘does not contain music’  
(although there is a famous piece of modern music  
that consists of five minutes of silence;   
but that is another matter…).  
  
6.3 V**ectors of standardized data.**

If, in addition to centering our variables,  
we also divide them by their standard deviation,  
we now have a vector standardized data  
(as shown in Excel sheet 7).  
  
Remember that we concluded above  
that after standardization   
all data series have equal standard deviation, equal to 1.00.  
  
If you standardize the height and weight observations, for example,   
they would both be represent as an arrow pointing in 8-dimensional space,  
with each arrows of equal length 1.00,  
and the angle between them representing the direction and intensity of their association.  
  
We have reached an important end-point!  
If you have a table of data,  
for example the table of height, weight and sports intensity data  
for eight children  
as in Excel sheet 7,  
you can picture this table as follows   
after standardizing the values of these variables:  
with each variable corresponds an arrow of length 1.00;   
the angle between each pair of arrows reflects   
the extent to which these variables, these measurements, are associated:  
- positively or negatively?   
(pointing in generally the same way or in opposite ways)   
- closely associated with each other or not?  
(sharp or not so sharp angle between each pair of variables).  
  
**6.4 The Correlation as a measure of association**.

Let us look at the data series or vectors for height and weight again.  
 Height: H’ = (112, 115, 112, 106, 120, 105, 111, 109)

Weight: W’ = (50, 51, 46, 49, 54, 51, 59, 55)

Is there a number that could express   
the direction and extent of association between them?  
Ideally, this would be a number that is positive if both concepts ‘co-vary’ positively  
and negative if the co-vary negatively,   
and zero if they do not co-vary.  
Further, it would be nice if such a number would be like an index,  
i.e. a number that lies between zero and one.

One way to express the extent to which the numbers for height and weight   
(and the concepts they measure) co-vary,   
would be to multiply out these numbers   
and to average the resulting sum of cross-products;  
i.e. the operation H’W/8 *(see Excel sheet 7).*   
  
(112\*50 + 115\*52 + 112\*47 + 106\*48 + 120\*55 + 105\*43 + 111\*48 + 109\*49)/8 = 5.464,5   
  
On closer inspection, this is not what we want:  
the sum of cross-products of the original numbers is always positive  
and the resulting average of the cross-products does not lie between zero and one.

Another approach would be to use   
the average of the sum of cross-products of the centered values.  
  
If above-average height generally goes together   
with above-average (or below-average) weight  
(and vice versa),  
then the average cross-product will be positive.  
If above-average height generally goes together   
with below-average (or below-average) weight  
(and vice versa),  
then the average cross-product will be negative.  
If above-average height does not systematically go together   
with above-average (or below-average) weight  
then the average cross-product will tend to be zero.  
  
Let us look at the centered values for height   
 Hc = 0.75 3.75 0.75 -5.25 8.75 -6.25 -.25 -2.25.  
and for weight  
 Wc = -1,875 -0,875, -5,875 -2,875 2,125 -0,875, 7,125, 3,125  
  
When we look at these two centered series,   
we see that they tend to co-vary positively:   
positive centered height tends to go together with positive centered weight  
and negative centered heights with negative centered values   
but not perfectly;   
large deviations in one variable tend to go together with large deviations in the other;   
we say ‘tend’ because, again, the tendency is not perfect.

If we compute the average cross-product of the centered values,   
(Hc’\*Wc)/8   
the result is called **the ‘covariance’** between both series or concepts;  
the covariance is seen to be 2.66;  
this number expresses something about how both series co-vary (here positively).

The number could have been negative (in the case of negative association)  
or even zero (no association).  
  
So, this ‘**covariance**’ seems to better meet our needs,  
but it still is not exactly what we want:  
the resulting number is not constrained to lie between zero and one.  
Even worse: the covariance number   
 depends on the measurement unit used,   
on the ‘yardstick’ we use to measure each of the concepts.  
If we had measured height in meters, rather than in centimeters, for example,   
the resulting covariance would be 0,0266.   
Yet, heights remain the same, irrespective of whether you measure them  
in meters, or centimeters, or feet, or whatever.

Could we not have a measure of association that does not depend   
on whether you measure height in meter or centimeter  
and weight in kg or in gram,   
a number a bit like a covariance,   
but not affected by the measurement unit chosen?  
  
The solution is to standardize the measurements,  
i.e. to use the standard deviation of each set of measurements as its measurement unit.  
Then, each measurement is ‘constrained’ to the same standard deviation,  
namely 1.00 as shown above.   
If you compute the average cross-product of the standardized values of two series,   
the result is called the **correlation coefficient**.  
  
In this example, the correlation coefficient   
between height and weight is seen to be 0,16 *(Excel sheet 7).*  
  
The correlation coefficient   
is an index of association between two phenomena, variables or measures   
it lies between -1 and +1.   
  
+1 indicates a perfectly positive association between both variables;   
(if you know the value of one, you can compute the value of the other exactly;   
they are then **linearly dependent, like °C and °F**; remember?)   
You would then be able, for example  
to predict exactly someone’s weight from his/her height   
(which we know is not possible; the correlation of height with weight will be less than 1.00).  
  
When the correlation is 0.00, the two variables are totally unrelated.  
They are also said to be ‘independent of one another’   
or ‘orthogonal’ to one another.  
That is not something we would expect between e.g. height and weight.   
  
When the correlation is -1.00,   
the two variables are related perfectly negatively;   
you could perfectly compute the value of one from the negative or reverse of the other;   
obviously, that is also not something one would expect between height and weight numbers.

With our height and weight example   
(remember: we have invented these data),  
the correlation is computed to be 0,16,  
a positive but weak association.

All of this reminds us of what we mentioned above,  
when discussing Pythagoras’ rule,  
and the representation of two variables or measurement series as vectors,  
with an angle between them,  
and the cosine of that angle expressing   
the direction and extent of the association  
as an index lying between -1.00 and +1.00.

**The conclusion is that the correlation coefficient   
between two variables or series of measurements  
is actually the same index   
as the cosine of the angle that lies between them in a graphical representation.  
If both variables overlap (angle of 0), their correlation is +1.00.  
If both variables are at right angles (‘orthogonal’, ‘independent of one another’),  
their correlation is 0.00.  
If both vectors lie in exactly opposite directions,   
their correlation is -1.00.**

**In real applications, the correlation between two variables  
will, of course, seldom be exactly 1.00 or -1.00 or exactly 0.00.  
The correlations coefficient will normally have a value somewhere  
between +1.00 and -1.00.  
In our example, for example, it is 0.16.**

**To compute the correlation coefficient between two series:  
standardize both series  
and compute their average cross-product.  
  
You must compute a correlation coefficient by hand or by Excel  
once (and only once!) in your life  
so as to understand how it is computed.  
  
Since the correlation is so basic to data analysis,  
all data analysis programs, including Excel,  
compute that coefficients when you request it,  
and sometimes even without your request.**

The correlation is one of the most important concepts  
in data analysis.  
It tells us whether, and how, two variables, measures or concepts  
are related to one another.  
But how about the situation when there are more than two variables?

**6.5 Multiple correlation**

When two variables are involved, we speak of a **simple correlation**.  
When more than two variables are involved,  
we speak of a **multiple correlation.  
The multiple correlation coefficient**is an index between -1 and +1 that expresses the extent to which   
a group of variables or measurements are related to each other,  
the extent to which these variables ‘share the same information’**.**

For example,  
if we also have a third measure about our eight children,   
let us say the extent to which they engage in sports,  
measured on a scale from 1 (not at all) to 5 (very much).  
*We have entered some data like this in Excel sheet 7  
(remember: just an example with invented numbers!).*

The weight of a child may reflect its height,   
but maybe also its sportiveness;   
we may expect a relationship between its weight on the one hand   
(represented by the symbol y)  
and its height (x) and sportiveness (z) on the other hand.  
  
This, written as a (linear) model, would be  
   
 y = ax + bz  
  
where the parameters a and b must be defined.  
   
If we have an idea about the value of a and b, e.g. a = 0,5 and b = 0,25,  
it is easy to compute the correlation coefficient  
between weight (y) on the one hand,   
and the ‘weighted sum’ ax + bz, on the other  
in this case 0,5x + 0,25z   
(we use the standardized value of these three variables to make things easy).  
  
We find the correlation between weight  
and 0,5\*height + 0,25\*age   
*in Excel sheet 8   
(using the instruction formulas>more functions>statistical>correl)*it is 0,16.  
  
But what if we do not know the values of a and b?  
Can we find a value for these two parameters that is ‘best’?  
And what do we mean by ‘best’ in this context?  
  
Best in this case may be: the values a° and b°   
which make the correlation between y and (ax + bz)   
as strong as possible.  
  
Again we use the least square principle:  
select values of a and b   
that minimize the sum of squared deviations  
  
 {y – (a°x + b°z)}2   
  
Again, we can instruct Excel   
to find these parameter values a° and b°

The resulting correlation is called a **multiple correlation**:  
it expresses the extent to which three or more variables are related.  
That multiple correlation can here be computed to be 0,25.  
  
With four variables, y, x, z and w, we may ask for the correlation  
between y on the one hand, and on the other hand x, z, and w;  
by this we mean the correlation between y   
and the ‘(linear) combination’ (ax + bz + cw).  
Again, this can be done according to the least squares principle.

While we are at it, we could also compute a correlation  
between one set of two variables (e.g. y and x)  
and another set of two variables (e.g. z and w);  
applying the least square principle.  
This means that we look for the values of the parameters a, b, c, and d   
which make the correlation between  
(ay + bx) on the one hand and   
(cz + dw) on the other hand most pronounced.

In other words, we look for values a°, b° ,c° and d°  
that minimize the sum of squared differences  
 {(a°y + b°x) – (c°z + d°w)}²   
  
That, again, is an exercise which Excel can be instructed to carry out for us.

1. **Data tables and matrices**

**7.1 Matrices and matrix operations.**  
In most empirical research,   
we observe or measure more than one variable  
and we do that for more than one ‘entity’.  
In the example above,   
we worked with the variables height, weight and sportiveness  
(3 variables)   
and we measured eight children  
(8 ‘entities’).  
  
The result is an ‘8-by-3’ of 8\*3 table of data,   
also called a data matrix, an 8\*3 matrix,  
with 8 rows and 3 columns,  
as in Excel sheet 8: rows 3 to 10 and columns B, C and D.  
  
Such data tables are typically constructed   
with rows corresponding to ‘entities’ and columns corresponding to variables.  
Entities could be people, countries, days, etc.   
  
We are quite familiar with data tables nowadays.   
Excel is one of the computer programs used most often  
to enter, show and to analyze data tables or sheets.  
   
If you interview 1500 people and ask them to answer 20 questions,  
the result would be a data table or matrix   
with 1500 rows and 20 columns, an 1500\*20 matrix,   
call it M.  
M would contain 1500\*20 = 30.000 codes (numbers or other symbols)  
representing the answer of one person to one question.  
  
If we represent a data matrix by the symbol **M,**for example the 3-by-2 matrix 2 4  
 1 5  
 3 1then the following operations on matrices can be usefully mentioned  
(and are available in Excel):

- matrix transposition  
like with a data vector, we can take the transpose of M, written as M’;  
if M is a (3-by-2) matrix, then M’ is a (2-by-3) matrix of numbers;  
M’ = 2 1 3  
 4 5 1  
  
- matrix multiplication  
if we multiply M’ with M, then M’\*M or M’M is a (2-by-2) table;  
you can multiply these two matrices, because they are ‘commensurate’,  
i.e. their inner dimension accords : M’ is a 2\*3 matrix and M a 3\*2 matrix  
the result of (2\*3) by (3\*2) is a 2-by-2 matrix.  
That matrix actually contains the sum of cross products off its diagona**l**and the sum of squares on its diagonal:  
  
M’M = (2² + 1² + 3²) (2\*4 + 1\*5 + 3\*1) = 14 16  
 (4\*2 + 5\*1 + 1\*3) (4² + 5² + 1²) 16 42  
  
- matrix inversion  
This is a rather obscure operation, but one that we can explain intuitively.  
The inverse of the number 3 is 1/3 or 3-1, because 3\*3-1 = 1.  
The inverse of a matrix   
(this is only possible for ‘square’ matrices, e.g. 2-by-2 or 3-by-3, etc.)  
is another matrix  
such that if you multiply the latter with the former   
(you multiply the inverse of the matrix with the matrix)  
the result will be a matrix with 1’s on the diagonal and 0’s off-diagonal.  
The latter matrix is called an ‘Identity’ matrix;   
that is because if you multiply a matrix M with an identity matrix I,  
then the result is the matrix M.  
  
For example, the inverse of the matrix M’M above 14 16  
 16 42  
is thematrix (M’M)-1 which can be computed (e.g. by Excel) to be   
 42/(14\*42-16²) -16/(14\*42-16²) = 42/332 -16/332  
  
 -16/(14\*42-16²) 14/(14\*42-16²) -16/332 14/332  
  
If you multiply out (M’M)-1 with (M’M), the result is the (2-by-2) matrix 1 0  
 0 1  
How matrices are inverted is beyond the purpose of this short course.  
But you understand the analogy with the inverse of simple numbers (i.e. 3\*3-1 = 1).  
Inverting matrices will be necessary later (in other course)   
when processing data tables for statistical analysis,   
especially when these tables become large.   
As mentioned, Excel can transpose, multiply and invert data matrices for you.

**7.2 Analysis of data tables (matrices)**

In research, data matrices are often very large.   
A questionnaire study with a sample of 1000 boys   
with some 60 questions about their knowledge, attitudes and practices in the domain of sexuality   
would result in a 1000\*60 data matrix   
(1000 boys and the answers to 60 questions, or 60 000 numbers…)   
  
Needless to say that we can easily get lost in a table of 60 000 numbers.  
  
We should therefore be happy to find ways to make such data tables   
easier to represent and understand.

This section deals with that problem;  
it is therefore very relevant for empirical researchers  
and will be used in many of the courses that follow this introduction.  
At the same time, many concepts that were introduced above  
will be applied here; this is your reward for having studied those topics!  
  
Depending on how you approach the (3\*8) data table,   
i.e.: from the side of its rows or from the side of its columns,  
you see (from the top down) three vectors of eight numbers each,   
corresponding to the columns of the matrix,  
e.g. three variables: height, weight and sportiness   
or (from the side, from left to right) eight vectors of three numbers each,   
corresponding to the rows of the matrix  
e.g. eight children.

We will show how to make a data matrix easier to represent or understand  
by approaching it from the side of its rows (or the entities) first   
and from the side of its columns (or variables) next.

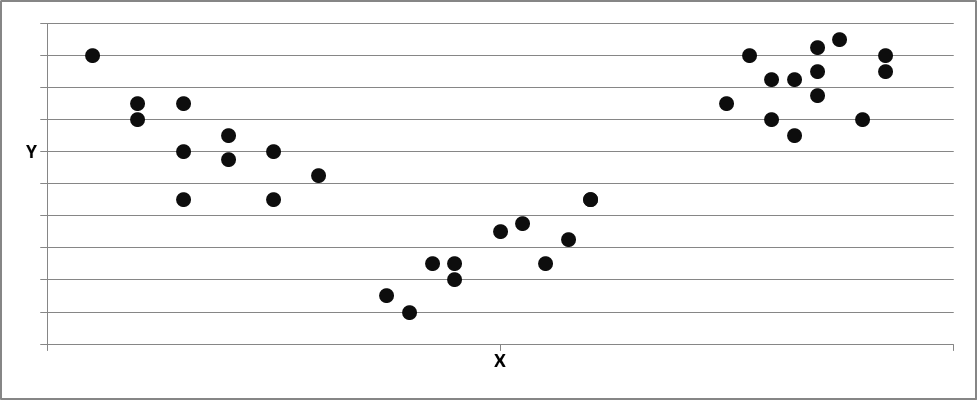
**7.2.1 Approaching a data table through the entities (rows).**  
  
First, let us approach the data table from the side of the rows,   
i.e. from the side of the entities (in our example: the 8 children).   
For each child, we have recorded three variables: height, weight and sportiness.   
Approach these three numbers as coordinates on three axes or dimensions,  
one for height, one for weight and one for age.   
Each boy is then represented as a point in (three-dimensional) space.   
  
Together, they form a ‘scatter’ of 8 points in the 3-dimensional space.   
In this case, the scatter is hardly visible, because it contains only 8 points;   
but if a much larger number of entities had been observed (e.g. 1000 children),   
the scatter would be denser and more clearly visible as a data ‘cloud’.  
  
Note that, if we had recorded a fourth variable,   
e.g. the IQ-score of the child,   
that would lead us to represent the child as a point in a four-dimensional space;   
as already discussed above, that is something that we cannot imagine visually,   
but that we can represent or ‘picture’ in our mathematical mind.  
And the same goes for a fifth, sixth, etc., variable (and dimension).   
  
Instead of using the original measures in our data matrix,   
we could also use the centered values and/or the standardized values.   
As explained above, there is generally a preference   
to work with standardized values of variables.

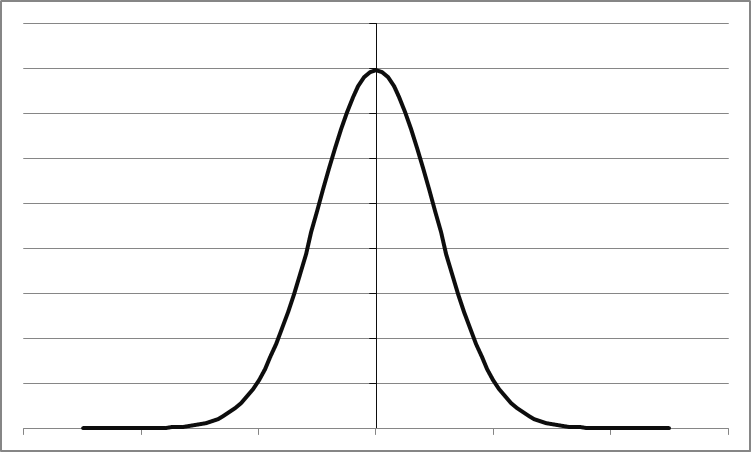
Upon inspection of the data cloud,   
we may be interested in the structural characteristics of that cloud:  
its structure and shape.  
  
- is there just one cloud, or are there several (sub)clouds?  
  
On the basis of the variables measured,   
is there only one, (rather) homogeneous group of entities,   
or can we discern  
more than one groups,   
that are internally homogeneous  
(i.e. observations belonging to the same ‘sub-cloud’ are very similar to one another)  
and are externally heterogeneous  
(i.e. observations belonging to different sub-clouds are very different from one another).

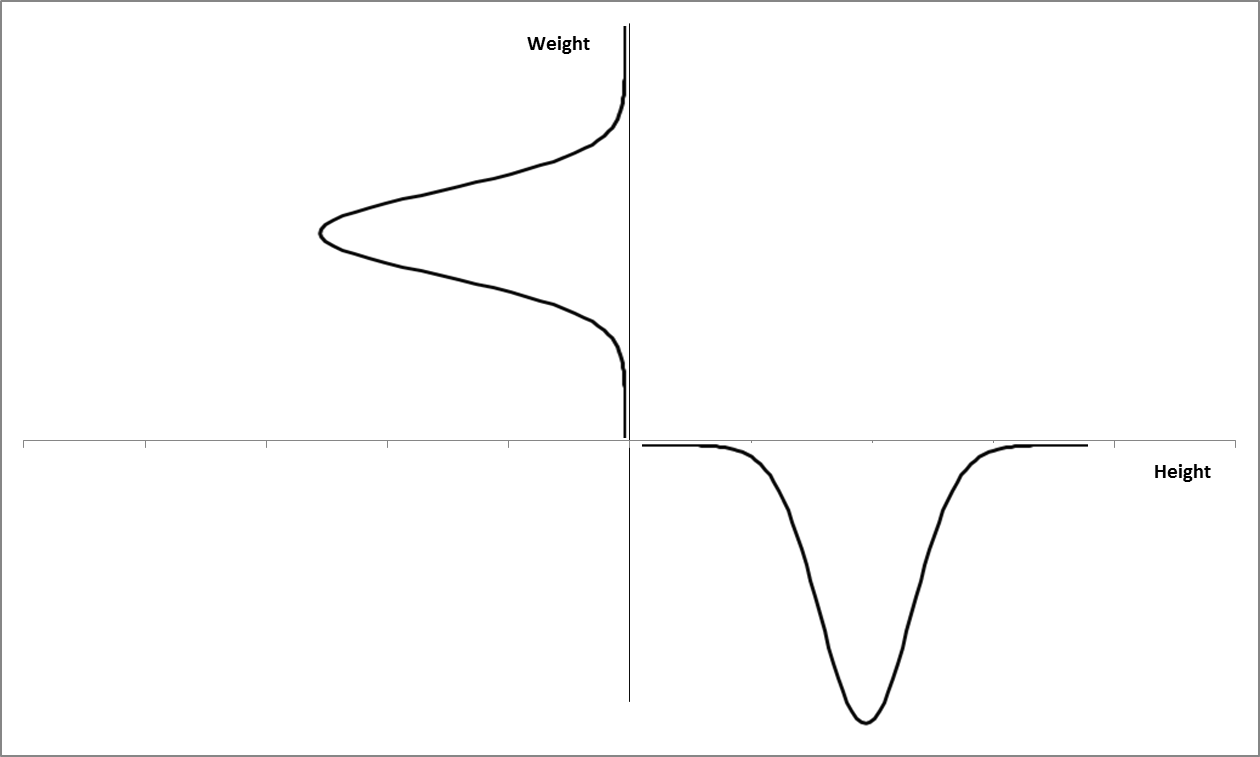
Such groups are called ‘clusters’ in statistics, or ‘segments’ in marketing,   
or types in other disciplines, etc.  
This is a question that we encounter in most of the empirically based disciplines.

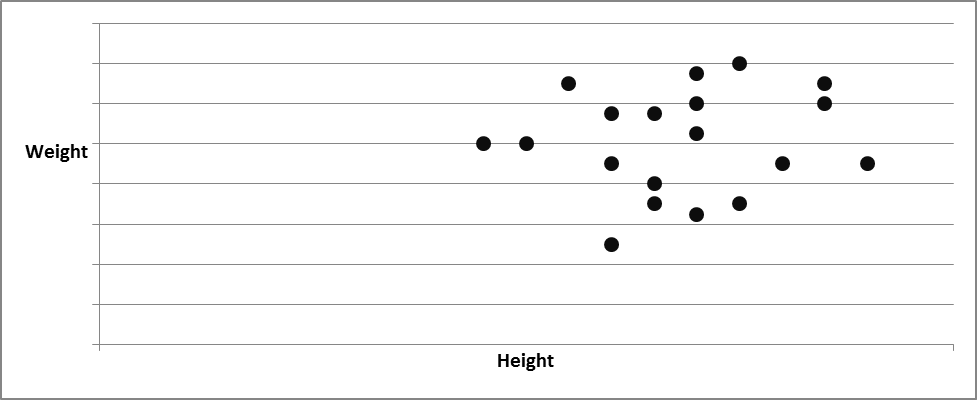
In the graphical example below,  
with a scatter diagram based on two variables,  
we seem to be able to discern three groups.  
With many hundreds or thousands of observations,  
and with many variables,  
it becomes difficult to discern such groups with the naked eye.

Determining the number of such clusters,   
identifying to which subgroup or cluster an entity belongs  
and describing the nature of each subgroup   
is the object of statistical techniques   
with names such as ‘**clustering**’ or ‘**discriminant analysis**’.



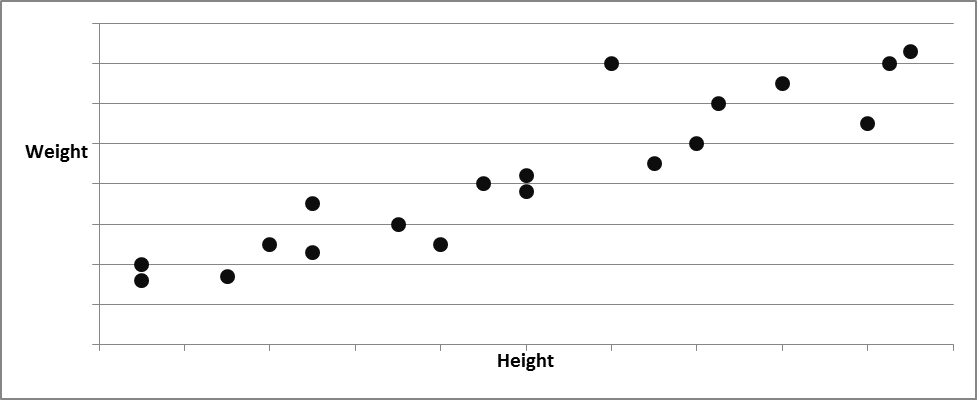
- do the data clouds have a specific shape?  
  
The question here is how the observations are distributed  
over the values that the concerned variable may take.  
   
A single characteristic of a group or population,   
e.g. the height or the weight,   
is often distributed according to a regular distribution function;  
in many cases that distribution is Normal or Gaussian,  
with a concentration of observations around the mean or average,  
and with fewer observations further above or below the mean  
  


With data tables,   
we are not looking at a single variable,   
but at several variables simultaneously.  
  
Consider, for example, variables such as the height and weight of the adult males.  
Each of these variables tends to follow its own Normal or Gaussian distribution.  
  


How would we expect the scatter diagram   
of points corresponding to the observed pairs (height, weight) for adult males  
to look?   
  
If both variables, height and weight, would be unrelated,  
then the scatter diagram would look like a circle or disk,  
with a more dense concentration of observations in the middle  
and with more scarcely spread observations towards its edge.  


The circular shape would indicate that both variables are unrelated  
(their correlation would be zero).

But we know that height and weight are rather strongly related:  
if you are tall, you tend to weigh more  
and if you are small, you tend to weigh less.  
In this case, the data scatter would look like an ellipse,  
one with positive inclination,  
expressing that there is a tendency for height and weight to co-vary positively.  
Again, you might expect the observations to be packed more densely  
towards the center of the ellipse than towards its edge.



Were the relationship between both variables negative,   
then the ellipse would have a negative inclination,  
expressing a negative correlation,  
as you would expect, for instance, between the age of a person and his/her reaction speed.  
  
As the two variables became more strongly associated,  
i.e. as their correlation became more pronounced,  
the ellipse representing the entities would become tighter,  
and ultimately collapse in one line or linear relationship   
when the correlation becomes 1.00 in value;  
in the latter case you can predict the value of one variable exactly  
if you knew the value of the other (i.e. linear dependence).

If we move from two to three variables,  
the data scatter is now spread in space;   
it will have a spherical shape   
(if all three variables were essentially unrelated)  
or the shape of a boulder (i.e. a ‘three-dimensional ellipse’)   
(if the variables have some relationship).  
The shape and orientation of the data scatter   
tells us something about the association (the correlation)  
between the three measures or variables.   
  
A spherical shape would indicate low correlation.   
A sharp ellipsoid (‘pebbly’) shape would indicate a pronounced correlation.

With four or more variables, the representation is essentially the same,  
only now, we cannot picture it visually;  
we have to appeal to our mathematical imagination  
to picture e.g. a six-dimensional ellipse, etc.   
  
But this is basically the same as when we work with a number of dimensions  
that we can visually imagine.

**7.2.2 Approaching a data table through the variables (columns).**

When measuring one variable for a number of entities one obtains a data vector.  
  
We repeat that which has been amply discussed above:   
if we standardize the measures, the corresponding vector   
can be pictured as an arrow of length 1 pointing   
from the origin in some direction.   
Two (standardized) variables with a correlation of value ‘r’  
can be represented as two arrows of identical length 1  
starting from the same origin  
and making an angle of v degrees,  
such that the value of cos(v) equals the value of the correlation coefficient ‘r’.

With these principles,  
if we know the value of the correlations between its variables,   
we can again represent a data matrix   
but now approached through its columns (or variables).

Computing the correlations between all variables in a data matrix  
can be shown to be easy:  
a correlation between two variables, after all,   
is the average cross-product of their standardized values.  
If you take the data matrix M and standardize the values of its variables,  
to obtain the matrix **Ms** of standardized values,  
then the standardized cross-products between the variables   
are obtained by matrix multiplication (**Ms’\*Ms**);  
divide this by the number of observations,  
and you obtain the matrix R of correlations  
(more precisely: R will show the correlations of the diagonal;  
the diagonal elements of R are 1.0;  
these are the variances of the standardized variables).   
  
As we said, no need to computer R yourself,   
Excel or other computer programs will do that for you.  
  
We can give a representation of the three variables in our data matrix:  
height is represented by means of an arrow of length one;  
weight can next be entered as a second arrow of length one,   
starting from the same origin  
and making an angle v with the arrow representing height,  
such that cos (v) equals the correlation between height and weight.  
That ‘picture’ will fit in a (flat) plane, the plane of the tow arrows.   
  
The variable ‘sportiness’ can now be shown as a third arrow,  
sticking out from the same origin,   
and making angles with height and weight  
corresponding to its correlations with these two variables.

Two variables (their arrows) can be represented in a plane (‘on a sheet of paper’);  
representing three variables will require a third dimension  
(need not only ‘North-South’ and ‘East-West’,   
but also ‘up-down’ to show the three arrows).  
It may be that the third arrow sticks out only very lightly   
from the plane of the other two.  
In that case the multiple correlation between sportiness on the one hand  
and weight and height on the other is high,  
that ‘sportiness’ shares a lot of its information with height and weight.  
  
Or it might be that the multiple correlation is close to zero;   
the vector for sportiness would then be rather orthogonal   
to the plane made by the other two arrows  
(it would clearly stick out in the ‘up-down’ direction),  
indicating that it has little in common with these two other variables.

Of course, a number of other situations between these two extremes are possible.  
The three variable are represented by a ‘bunch’ of arrows;   
this bunch may have some structure,   
e.g. all three arrows stick close together,   
or two stick together while the third points in another direction,  
or all three point in a different direction, etc.  
  
This structure tells us something about which variables belong together  
(e.g. because they tend to measure the same thing,  
or because one causes the other,  
or because there is a background variable that causes both).  
   
If we would have still more variables,   
we simply extend what we did with three variables;   
only now we cannot represent this visually,   
as we need to portray a bunch of arrows in four, or five  
or even more dimensions.  
Again, try to imagine this with your mathematical imagination.

If we have drawn all the variables in our data matrix as arrows,  
with angles reflecting their mutual correlations,  
we are ‘looking’ at a bunch of arrows, a bit like a bunch of flowers.

We are interested in the eventual structure of this bunch of arrows.  
Two questions are of interest here:  
  
- how many dimensions are really needed to fit this bunch of variables  
This deals with the ‘dimensionality’ of the variables in our data table:  
how many dimensions are really involved in our variables?  
  
As just mentioned in the example:  
even though the three arrows stick out in three-dimensional space,  
they may lie ‘almost’ in two dimensions;  
then two dimensions might suffice to capture most of the data.  
  
If you ask 60 different questions in a questionnaire study,   
it is quite unlikely that these answers reflect 60 totally different concepts.  
There will probably be overlaps in information  
between so many variables.  
A bunch of 60 arrows might be captured by a much lower number  
of more general or more fundamental ‘underlying’ variables or dimensions,  
these are sometimes called **components or factors**.  
  
- are there groups of variables that ‘cling’ together?  
  
Some variables may belong together more with one another   
than with other ones:  
some variables my ‘cluster’.  
Graphically, you might imagine a bunch of arrows  
in which you can identify specific sub-bunches;  
the latter would then indicate subgroups of variables  
which are mutually more related than with other variables.   
Identifying how many such clusters are present,   
and what variables cluster together,   
again is a way to obtain a better view on the structure of a data matrix.

1. **Null hypothesis and alternative hypothesis; probability distribution.**

The final topic to be introduced and illustrated   
deals with statistical reasoning,  
with uncertainty and probabilities  
and making decisions when uncertainty prevails.  
This we do by means of an example with a fair die;  
but, as usual, the example is only meant to introduce  
a more general issue.

We introduced the concept of a model above.   
A model is a representation of a phenomenon,   
or of a relationship between several phenomena.   
We gave as an example   
the distribution of weights in an adult population,   
or the relationship between pushing the gas pedal and the speed of the car.  
  
These models originate from our own mind:   
they reflect how we represent the outside world in our mind.   
To some extent, therefore, these models are not necessarily correct;   
they just reflect how we think that the outside word is;   
in scientific language, we say that they are ‘hypotheses’   
(from the Greek ‘hypo’ and ‘thesis’, meaning ‘underlying assumption’).  
  
One popular view is that we can never know how the ‘outside world’ is,   
but that we only can make guesses about it;   
a further idea is that we can never prove that our guess is right,   
but that all we can prove our current guess to be wrong.   
As long as it not proven wrong, it is OK to keep using that guess  
as a working hypothesis.  
   
Once it is proven wrong, we replace it by a new guess, a new hypothesis;   
often this is an improved version of the old, now rejected, hypothesis.  
Occasionally, but seldom, we replace it by a totally new, different hypothesis.   
  
In science, and where the model involves relationships of the kind ‘A influences B’,   
one often works with a ‘null hypothesis’.   
The null hypothesis corresponds to the idea that ‘there is nothing going on’.   
If we are interested in the possibility of an influence of A on B,   
then our null hypothesis would be that A does not influence B.   
The scientist then usually sets out to prove that the null hypothesis,   
i.e. that A does not influence B,   
must be rejected in the light of the facts.   
The scientist may, for example,   
set up an experiment in which he manipulates A   
and demonstrates beyond reasonable doubt   
that B responds to that manipulation.  
If the null hypothesis is then not rejected by the facts,   
it is maintained (B is not influenced by A);   
if B responds to A, the null hypothesis is rejected   
in favor of the (an) alternative.   
  
Note that we never say that a hypothesis is proven right;   
we can only say that a hypothesis is proven wrong.  
This is a rather profound statement.   
It implies that our knowledge does not progress   
by demonstrating that our thoughts are ‘right’,   
but only by eliminating those of our ideas   
that are proven wrong by the facts.  
  
We used the expression ‘beyond reasonable doubt’ on purpose,   
because it is not always very easy to verify that our model is wrong,   
that a null hypothesis must be rejected.  
  
Let me give an example that you can understand,   
and which will be of use later as well.  
  
Let us assume that you have a die.   
A fair die is one where each of the six sides has an equal chance of being thrown.   
Now you challenge a friend to play a simple game,   
where he can win or lose 100$ depending on the outcome of the throw of the die.   
Your friend might ask you if the die is really fair, ‘unbiased’;   
in other words, he might be suspicious that the die is loaded   
(that is has a higher probability of landing on one side than on another,  
in which case it would not be ‘fair’).  
  
You agree with your friend to carry out an experiment,   
namely to throw the die a large number of times, e.g. 120 times   
and to see if the outcome is fair.   
  
Your **null hypothesis** here, would be that the die is fair,   
that each side has an equal probability of being thrown;   
the alternative hypothesis would be that some sides   
have a higher probability of being thrown that others.   
  
Let us say that we carry out the experiment,   
and note the following outcome

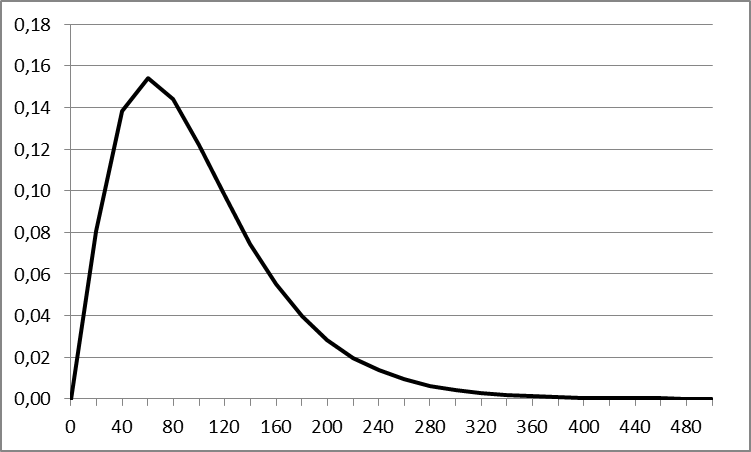
side n° of throws  
1 18  
2 23  
3 25  
4 20  
5 17  
6 17

Total 120

Observing this outcome, your friend might say:   
if the die is fair, I would expect it   
to land the same number of times on each side, i.e. 20 times   
(that is the null hypothesis: there is nothing there,  
there in no unfairness).  
He points out that the throws are not each 20,  
and concludes that the die is not fair.  
  
To this we would reply that he should take into account   
that there is uncertainty involved;   
if you do this experiment, you will never see   
exactly 20 throw for each side.   
What would matter is that the numbers show   
that there is a tendency towards 20 or each side of the die;   
that the figures do not depart too strongly from 20 throws each.  
that if we would throw the die not 120 times,   
but infinitely many times,   
we would find that each side is thrown   
with the same percentage, namely 1/6.   
  
The question therefore is whether the deviations from the null hypothesis   
in the numbers above are sufficiently strong   
that we can safely conclude that the die is not fair.  
  
When we look at the deviation of the outcomes   
from the expected outcomes (20),   
we see that there are some deviations are larger,   
and others smaller.   
It would not be wise to rely on a single one of the 6 numbers above   
to reach a conclusion.   
We must take into account the full profile of the results,   
of all the deviations or differences from the null hypothesis.   
  
What single number would be a good indication of the extent  
to which these results differ from the null hypothesis?  
  
We could just add up these differences,   
but as there as many positive deviations as negative ones,   
the result would always total zero,   
hence not reflect how extreme these deviations are.   
  
We have already seen above, that in such cases,   
it is often a good idea to square the deviations and to compute a ‘sum of squares’.   
  
This we did in the table below;   
the ‘sum of squares’ here is 56.   
  
We understand that the larger are the deviations   
from the expected number of 20,   
the larger will be the resulting sum of squares.  
  
The question now is whether 56 is such a large number,  
that it reflects such sharp deviations from the pattern expected under the null hypothesis,   
that we would have to discard the null hypothesis that our die is fair  
in favor of the (‘alternative’) hypotheses that it is ‘loaded’.  
  
side n° of throws expected n° of throws deviation deviation²   
1 18 20 -2 4   
2 23 20 3 9   
3 25 20 5 25   
4 20 20 0 0   
5 17 20 -3 9   
6 17 20 -3 9

Total 120 120 0 56

To see how we answer that question,  
let us carry out the follow ‘experiment’ in our mind (**a ‘thought experiment’**)  
  
We instruct the computer to draw 120 random numbers from one to six,   
with equal probability for each of the numbers 1, 2, 3, 4, 5, 6.   
In other words, we ask the computer to ‘throw a fair die’ 120 times;  
we have shown above how that can be done.   
and we note the sum of squared deviations that results from those 120 throws:  
we note the number of times our ‘computerized die’ lands on each side  
and note the sum of squared deviations.  
  
Now, we instruct the computer to do this (120 throws)   
a large number of times, e.g. 100.000 times,   
and each time we note the sum of squares.   
  
Sometimes the resulting number will be smaller,   
sometimes it will be larger,   
but always that will be due only to chance,   
since the computer’s die is fair   
(i.e. each side has a probability of exactly 1/6 of being thrown).   
  
We ask the computer to make us a diagram of the resulting values  
of the sum of squared deviations.   
That is only a small effort for the computer.

The resulting distribution will look as follows  
(along the x axis we show the observed values of the sum of squared deviations,  
along the y-axis, we show how often these values occurred).  
  


From this picture, we would conclude   
that the value of the sum of squares  
that we observed by throwing the real die, 56,   
is not at all extreme.   
Indeed, even a fair die would,  
just as the result of chance,   
produce a sum-of-squares value for 120 throws that exceeds 56  
more than about 73% of the time.   
  
The value of 56 for the sum-of-squares is therefore quite normal  
under the null hypothesis   
and indicates that our die seems to be fair.   
  
**We cannot reject the (null) hypothesis that the die is a fair one.   
It is therefore rational to continue under the assumption that the die is fair**.

The results could have been different.   
Let us assume that they were as follows:

side n° of throws expected n° of throws deviation deviation²   
1 10 20 -10 100   
2 23 20 3 9   
3 33 20 13 169   
4 20 20 0 0   
5 17 20 -3 9   
6 17 20 -3 9   
  
Total 120 120 0 296

The sum of squares is now much larger, as we would expect.   
Considering the distribution drawn by the computer   
for the sum-of-squares produced by a fair die,   
we would find that a fair die would lead to such an extreme value   
of the sum of squares only very exceptionally, namely only 1% of time.   
This would lead us to surmise that, maybe, the die is not fair.   
  
Considering this evidence,   
it may then be wiser to reject the null hypothesis that the die is fair.

To resume, the reasoning in statistics is as follows:  
many phenomena are characterized by uncertainty.  
If we have a hypothesis about the world that we want to check,  
it will therefore seldom be enough to just look at the facts   
to be able to conclude with certainty that our hypothesis is right or wrong.

What we need to do is to ask ourselves how the world would look  
under the null hypothesis, taking into account the uncertainty   
that besets many phenomena.  
We then compare the situation that we observe (the facts)  
with the distribution of possible outcomes that we expect to observe  
if the null hypothesis ,were true.  
  
We can then make a statement about the probability  
that the facts we observe (ore even more extreme facts)  
will occur under the null hypothesis.

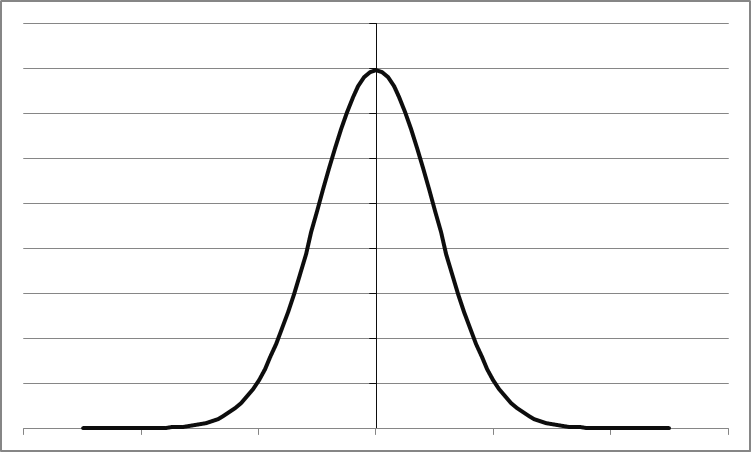
Fortunately, we do not generally have to follow the procedure  
that we used above with our die,  
instructing the computer to roll a die 120 times   
and repeat that 100 000 times  
to obtain the probability distribution   
of the sum of squared deviations under the null hypothesis.  
  
The reason for that is that  
1. most of these probability distributions can be derived mathematically  
2. the probability distribution of the outcomes of many phenomena   
under the null hypothesis tends to follow a limited number of patterns.  
  
Again, a law of nature seems to be at work here to make   
difficult things a bit more manageable…

Two probability distributions that tend to occur very often are  
the Normal or Gaussian distribution  
and the Chi-square distribution.

Let us take the last one first,   
because it is essentially the same as the probability distribution   
of the die example we gave above.

If we study a phenomenon which results   
in the occurrence of a limited number of categories,   
for example the results of 120 throws of a die   
(120 ‘events’, 6 possible categories),  
or the distribution of 742 registered purchases of a book   
over the 6 workdays of a week  
(742 events, 6 possible categories)  
or the gender (male, female) of 2.000 visitors to a trade fair  
(20.000 events, 2 possible categories),  
the chi-square number is always computed as follows:  
  
1. Count the actual number of events in each category  
2. Determine the number of events expected in each category according to your hypothesis  
3. Take the square of the difference between these two numbers   
and divide that by the expected number of occurrences under you hypothesis:  
(actual number –expected number )²/expected  
4. Add up all these numbers over the categories  
5. Look up the probability of that number in Excel or in tables of Chi-square distributions;  
you will have to indicate the ‘degrees of freedom’;   
that is just the number of categories minus one  
(do not ask why, that is for other courses…).  
   
The number that Excel will give you is the probability that a value   
as extreme or more extreme than your number   
can occur if you null hypothesis is true.

If, for example, there are 2.000 visitors to you trade fair,  
and your hypothesis is that your fair attracts ¼ women and ¾ men,  
i.e. you expect 500 women and 1500 men for a total of 2.000 visitors.  
Now you actually count 460 women and 1540 men.  
The chi-square value is (460-500)²/500 + (1540-1500)²/1500 = 3,2 + 1,07 = 4,27.  
  
Excel will tell you that the probability of a number so high, or even higher  
under the null hypothesis of ¼ women, ¾ men,   
is 2,3% : a rather low number.  
Such a result has only a probability of 2,3% if your hypothesis is correct.  
This would lead you to conclude that your hypothesis is probably wrong  
you would tend to reject the null hypothesis.

Another distribution often encountered in nature and science is the Normal distribution.  
   


The weight of adult males, for example, will tend to be distributed   
like in the picture above, around the mean or average weight,  
and with a given standard deviation or typical difference.  
And many other phenomena are distributed in a similar way.  
  
As already mentioned, the nature of this distribution is that  
fixed percentages of adult male weights (or whatever it is that you study)  
will lie within specific intervals around the mean of the weight of adult males,   
namely:  
68% of the weights will lie within (plus or minus) one standard deviation of the mean   
95% of the weights will lie within (plus or minus) two standard deviation of the mean  
99,7% of the weights will lie within (plus or minus) three standard deviation of the mean.

Again, this allows testing some hypotheses, or making some decisions.  
Assume that your industrial bakery produces cakes.  
The production has been set up to produce cakes of 200 grams.  
Of course, your cakes do not always weigh exactly 200 grams:   
you have weighed 1000 of the cakes,   
and you find that the weight of the cakes is normally distributed   
with a standard deviation equal to 5 grams.  
  
Now you want to set up the production line   
so that you produce less than 1,5 cakes in 1000   
of less than 200 grams.  
  
For what average weight of cakes should you set up your production line?

Answer:  
if you set up your line to produce cakes of 215 grams,  
then 99,7% of your cakes will lie in the interval of   
215 grams plus or minus 3 standard deviations of 5 grams,  
i.e. between 200 grams and 230 grams.  
0,3% of the cakes will have a weight outside that range,  
of which half (1,5 in 1000) will be below 200 grams.

Now, if yesterday your production line   
produced 2,5% cakes of less than 200 grams,  
could it be that your line is set up for a different average weight than 215 grams?  
  
Answer:   
that is quite likely!  
According to the numbers given above,   
5% of the observations (cakes) of a normal distribution will lie outside   
the range of the average plus or minus two standard deviations.  
That is 2,5% below the range and 2,5% above the range.  
The mean of the cakes of the production line   
is therefore likely to lie at 200 gram plus 2 standard deviations, i.e. at 210 grams.  
It is therefore best to reject the null hypothesis that the production line  
is set up to produce cakes of 215 grams on average.   
You will have to re-adjust your production line!