Quantitative Tools - Level 1 - Fall 2015
Louis de Charsonville

Lecture 1

Changes and Percentages

1 Introduction

Data is ever-present. Every second, a bunch of 30'000 GB of data is released and collected across the world. Newspapers, TV shows, radio podcasts are delivering and commenting data. Economics, sociology and others social sciences are primarily based on data. Data is numbers, facts, figures, information, evidence and details. The goal of this course is to give you the toolbox for reading, understanding, analysing and mastering data. To do so, we go through some of the basic concepts of statistics. Quantitative Tools is about descriptive statistic with a pinch of financial calculus and a favour of applied arithmetic.

What are Statistics ? Statistics are the practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.

The fairy tale of Statistics - a quick history of Statistics

Statistics are born with the will of collecting taxes and drafting. The first recorded uses of Statistics are census, dating back from 3000 BC, in Mesopotamia and Ancient Egypt. China empire, 2000 BC, also used Statistics for farming conversion tables and census. During the Roman Empire, Statistics became more prevalent as the *censor*, a Roman magistrate in charge of the Census - a register of the citizens and of their property, was very influential. The use of Statistics declined somewhat during Middle Age. However, one can note the census made under Charlemagne reign and the fire census made in France in 1328 (*State of fire of 1328*) by Philippe VI, which counted 61,098 fires (the measure of households at the times) in Paris. In 1526, the French economist Jean Bodin published *La République*, a book in which he pointed out the benefits of counts for policy-making. Hermann Conring, a German intellectual, was the first to lecture on Statistics in a course named *Noticia rerum publicarum*. During the XVII century, the use of Statistics expanded in England where William Petty devised techniques for land surveying and developed the concept of "political arithmetic" : "*the art of reasoning by figures, upon things relating to government*".

Statistics changed radically with the development of probabilities as done by Pascal works (1623-1662), Fermat (1601-1665), Huygens (1629-1695) and Bernouilli (1655-1705). In particular, Gauss (1777-1855) legacy (Gaussian distribution, Gaussian-Markov law) is a cornerstone for the use of advanced mathematics tools in statistics use.

The XX$^{\text{th}}$ century is characterised by the rise of mathematical statistics with the work made by Pearson (1857-1936), and Russians Markov (1856-1922) and Tchebychev (1821-1894) after whom is named a famous inequality used for the proof of the Law of large numbers[1].

Today, Statistics are widely used for both analyzing data (descriptive statistics) and forecasting (inductive statistics).

2 Preliminary vocabulary of Statistics

Like any other sciences, Statistics has its own vocabulary. Here is the short-list of some of the most widely used words and their definitions.

[1] The law of large numbers states that the sample average converges in probability towards the expected value, in other words that as a sample size grows, its mean will get closer and closer to the average of the whole population

- **Population** : The entire set of items from which the analysed data can be selected. Items from a population share at least one property in common.
- **Sample** : A subset of a population that is obtained through some process, possibly random selection or selection based on a certain set of criteria, for the purposes of investigating the properties of the underlying parent population.
- **Variable** : A quantity that varies and take different values. For example, the income of inhabitants of a given country is a variable. A variable can be either
  - **qualitative** : if the population is the students of a class, a qualitative variable is for instance the hair color
  - **quantitative** (also named numeric), the height for example.

  If the variable is *quantitative*, it can be either :
  - **discrete** or countable : for example, the number of children per family.
  - **continuous** or unnumbered. For example: the height

  Note that a continuous variable can be discretize by putting its values in intervals.

  A *qualitative* variable can be either :
  - **nominal** or categorical and represents categories (age, gender, ethnicity, etc.)
  - **ordinal** : incorporates a ranking such as rankings for the best scorers in the Premier League.

  Example 2.1. Qualitative vs Quantative Variables

Table 1: Example of a qualitative and a quantitative variable

	Hair Color	Height (cm)
Alice	dark-haired	174
Bob	blond-haired	180
Craig	dark-haired	176
Dave	red-haired	164
Eve	light-auburn	186
Franck	light-brown	173

Table 2: Discretization of height

	Height (by intervals)
Alice	165-175cm
Bob	175-185cm
Craig	175-185cm
Dave	<165cm
Eve	>185cm
Franck	165-175cm

- **Frequency** : The number of times that a particular value occurs as an observation (in an experiment). Frequencies are usually graphically represented in histograms (for quantitative data) or bar plot (for qualitative data).

  Example 2.2. The number of M&Ms in a standard 100g bag

	Frequencies of candies per bag (approx.)
Blue	10
Brown	7
Green	10
Orange	11
Red	8
Yellow	8

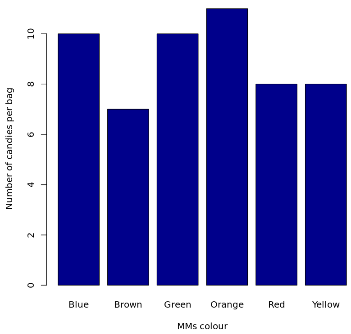

Figure 1: Bar plot of MMs per bag

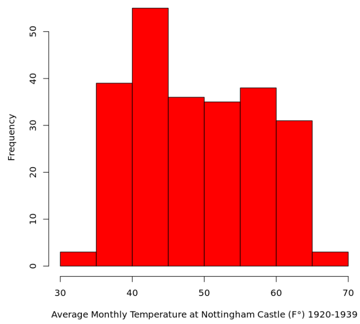

Figure 2: Average Monthly Temperatures at Nottingham, 1920–1939

- **Relative Frequency**: Relative frequency is another term for proportion; it is the value calculated by dividing the number of times an event occurs by the total number of events. The relative frequency is usually represented in percentages.

Example 2.3. The number of M&Ms in a standard 100g bag

	Relative frequency of candies per bag (approx.)	
Blue	0.19	19%
Brown	0.12	12%
Green	0.19	19%
Orange	0.2	20%
Red	0.15	15%
Yellow	0.15	15%

3 Absolute and relative changes, Percentages

In this section, we analyze two different measures of the changes of a variable : absolute and relative changes (another word is variation).

3.1 Definitions

3.1.1 Absolute change

The absolute change of a variable is the difference between its new value and its old value. For instance, if Alice's income was $40,000 in 2014 and is $45'000 in 2015. The absolute change of Alice's income is :

$$\$45,000 - \$40,0000 = \$5,000$$

Usually, we the greek letter Δ (delta) is used to denote the absolute change. So, if Alice's income is denoted by letter W, the absolute change of Alice's income between 2015 and 2014 is :

$$\Delta W = W_{2015} - W_{2014}$$

The absolute change of a variable should not be mistaken with the *absolute difference* which is the absolute value of the difference of two numbers (more info here : absolute difference - Wikipedia).

3.1.2 Relative change

The relative change of a variable express the absolute change relatively to the past value of the variable. The relative change of Alice's income between 2014 and 2015 is :

$$\frac{\$45,000 - \$40,0000}{\$40,0000} = 0.125$$

Formally written :

$$Relative\ change\ of\ X = \frac{\Delta X}{X}$$

So for Alice's income :

$$\frac{\Delta W}{W} = \frac{W_{2015} - W_{2014}}{W_{2014}}$$

The relative change is more useful than the absolute change when it comes to comparison : if the price of a brand new Tesla car goes from $87,500 to $87,501 and the price of the cream cheese bagel down the street goes from $2 to $3, the absolute change is exactly the same in both cases but the comparison does not make a lot of sense. In contrast, the relative change of the Tesla's price is 0.00001 while cream cheese bagel's is 0.5 (that's 43,750 times less).
The relative change is equivalent to the absolute change stated in the *variable unit*. For example, one could say that the price of the bagel has increased by 0.5 'bagel price'.

Relative change is usually stated in percentages. Thus, one would rather say that's the price of the bagel has increased by 50% (and the price of the Tesla by 0.001%).

3.1.3 Percentages

An old and widespread convention is to express *rati*o or *relative* change as a fraction of 100.
For example, 344 women out of 470 on board Titanic HMS survived the sinking ; so 73% of the women survived :

$$\frac{344}{470} * 100 = 73\%$$

In percentages, relative change becomes :

$$\boxed{Relative\ change\ of\ X = \frac{\Delta X}{X} * 100 = \frac{X_{final} - X_{initial}}{X_{initial}} * 100}$$

3.2 Operations on percentages

3.2.1 Multiplying factor

When doing calculus on percentages to describe relatives changes, it's better to use the multiplying factor. For a relative change of $x\%$, **multiplying factor** equals to

$$\boxed{1 + \frac{x}{100}}$$

Using the multiplicative factor, you won't make the error of thinking that an increase by $a\%$ followed by a decrease by $a\%$ is a return to the initial value (see below, **Example 3.2**).

Example 3.1. Bob's income between 2013 and 2015
If Bob's income has increased by 25% in 2014 and decreased by 20% in 2015, how has changed Bob income between 2013 and 2015 ?
Imagine Bob was paid $100 in 2013. As his income increased by 25%, he was then paid in 2014 :

$$\$100 + \frac{25}{100} * \$100 = \$125$$

And as his income decreased by 20% in 2015, his income is in 2015 :

$$\$125 - \frac{20}{100} * \$125 = \$100$$

So Bob is exactly making the same amount of money in 2015 as he was in 2013.

Using the multiplying factor:
Bob's income was multiplied by 1.25 in 2014. In 2015, Bob's income was multiplied by :

$$1 + \frac{-20}{100} = 0.80$$

So, Bob's income between 2013 and 2015 was multiplied by $1.25 * 0.8 = 1$

Hike and drop : no symmetry : A hike of $x\%$, succeeded by a drop of $x\%$ is not a return to the initial value.
Indeed :

$$y * (1 + x) * (1 - x) = y * (1 - x + x - x^2) = y * (1 - x^2) \neq y$$

Example 3.2. The books of Mr. Swann
Charles Swann, a book collector, has 10,000 books. After a walk to Combray Flea Market, he increased his collection by 15%. But unfortunately, a fire destroyed 15% of his collection the following night.
Does Mr. Swann have more or less books after the fire than he did before his walk to Combray Flea Market ?
Less.
Recall that Mr. Swann had 10,000 books before his walk to the flea market. After his purchases, Mr. Swann has :

$$10,000 * (1 + \frac{15}{100}) = 1.15 * 10,000 = 11,500$$

And after the fire which destroyed 15% of its collection :

$$11,500 * N * (1 - \frac{15}{100}) = 1.15 * 0.85 * 11,500 = 9775$$

At the end, Mr. Swann has only 9775 books left.

Successive changes are independent of their order of application : An increase by $a\%$ followed by a increase of $b\%$ is the same as an increase by $b\%$ followed by an increase of $a\%$.

Aggregate change of successive changes : An increase by $a\%$ followed by a increase of $b\%$ equals to an increase of :

$$y * (1 + \frac{a}{100}) * (1 + \frac{b}{100}) = y * (1 + \frac{a}{100} + \frac{b}{100} + \frac{ab}{100 * 100})$$

Example 3.3. The factory of Mr. Wonka
The annual output of the factory of Mr. Wonka is 1 million kilograms of chocolate. Thanks to a groundbreaking process, the output of the factory increase by 10% the first year and by 15% the second year. In two years, the output of Mr. Wonka has increased by 26.5% (and not 25%):

$$(1 + \frac{10}{100}) * (1 + \frac{15}{100}) = 1 + \frac{26.5}{100}$$

3.2.2 Average growth rate

In the following sentence :

"*For the past two years, the output of Mr. Wonka's factory has increased on average by* 12%"

One means that *if the growth rate has been (each year) the same during the last two years*, the output would have grown each year by 12%. Those 12% is the average growth rate.
How can one find the average growth rate ?
We know that the first year the output has increased by 10% and by 15% the second year. The aggregate change is the product of the two multiplying factors for the two years, so :

$$\left(1 + \frac{10}{100}\right) * \left(1 + \frac{15}{100}\right) = 1 + \frac{26.5}{100}$$

Let the initial value of the output be $Y_{initial}$, the final value of the output Y_{final}. We have :

$$Y_{final} = Y_{initial} * \overbrace{(1 + \frac{10}{100})}^{\text{increase by 10\% the } 1^{st} \text{ year}} * \underbrace{(1 + \frac{15}{100})}_{\text{increase by 15\% the } 2^{nd} \text{ year}}$$

$$Y_{final} = Y_{initial} * \left(1 + \frac{26.5}{100}\right) \quad (1)$$

Let now $x\%$ be the average growth rate. Because the average growth rate is the annual growth rate if the growth has been the same each year, it means that each year the output rate would have grown by $x\%$. If we use the aggregate change formula, we have :

$$Y_{final} = Y_{initial} * \left(1 + \frac{x}{100}\right) * \left(1 + \frac{x}{100}\right)$$

$$Y_{final} = Y_{initial} * \left(1 + \frac{x}{100}\right)^2 \quad (2)$$

Using now the formula (1) and comparing the aggregate change if the two situations (the real one and the one if the growth rate had been the same each year) we have :

$$\left(1 + \frac{x}{100}\right)^2 = \left(1 + \frac{26.5}{100}\right)$$

So, using the $\sqrt{}$ on the left-hand side of the formula we have :

$$\left(1 + \frac{x}{100}\right) = \sqrt{\left(1 + \frac{26.5}{100}\right)}$$

Finally, we have :

$$\frac{x}{100} = \sqrt{\left(1 + \frac{26.5}{100}\right)} - 1 \approx 0.12$$

So the average growth rate of the output is around 12%.

Formally:
Let Y a variable that grows at each period n (a period can be a month, a year) at the growth rate x_n. We note Y_{n+1} the value of Y at the end of period n. So we have :

$$Y_{n+1} = Y_n * (1 + x_n)$$

Let N be the total periods, so the n is taking the value from 0 to N, with Y_0 the initial value of Y and Y_N the final value of Y. We have :

$$Y_N = Y_0 * (1 + x_1) * (1 + x_2) * ... * (1 + x_N)$$

The average growth rate is the growth rate that would have take you from initial value Y_0 to final value Y_N, while being the same rate at each period.
Let z be the average growth rate. Since the average growth rate is the growth rate for each period n if the growth rate had been the same during the total number of periods N, then each year the multiplying factor would have been $(1 + z)$ [2]. So we have :

$$Y_N = Y_0 * \underbrace{(1 + z) * (1 + z) * ... * (1 + z)}_{\text{N times}}$$

So :

$$Y_N = Y_0 * (1 + z)^N$$

Using the root we have :

$$1 + z = ((1 + x_1) * (1 + x_2) * ... * (1 + x_N))^{\frac{1}{N}}$$

So the aggregate change is :

$$\boxed{z = ((1 + x_1) * (1 + x_2) * ... * (1 + x_N))^{\frac{1}{N}} - 1}$$

[2]note that here z is not stated in percentages. But if it was, then the multiplying factor would have been $1 + \frac{z}{100}$, if you are not sure refer to the example below

Example 3.4. The factory of Mr. Wonka (ctned)
We have the following problem : The output of the factory increased by 10% the first year, 20% the second year, 15% the third year and decreased by 15% the fourth year. What is the average growth rate on the whole period ?
Here :

- $N = 4$ is the number of years
- Y the output of the factory and Y_0 the initial value of the output, Y_1 ,Y_2 ,Y_3 ,Y_4 the outputs of the factory at the end of each year.
- $x_1 = \frac{10}{100}$, $x_2 = \frac{20}{100}$, $x_3 = \frac{15}{100}$, $x_4 = \frac{-15}{100}$ are the growth rates for each year.

We have :

$$Y_4 = Y_0 * (1 + \frac{10}{100}) * (1 + \frac{20}{100}) * (1 + \frac{15}{100}) * (1 + \frac{-15}{100})$$

If the growth rate had been the same during the four years, with the same initial and final value, and z that constant annual growth rate (z is also *by definition* the average growth rate), we would have :

$$Y_4 = Y_0 * (1 + z)^4$$

So we can deduce :

$$(1 + z)^4 = (1 + \frac{10}{100}) * (1 + \frac{20}{100}) * (1 + \frac{15}{100}) * (1 + \frac{-15}{100})$$

And finally :

$$z = \left((1 + \frac{10}{100}) * (1 + \frac{20}{100}) * (1 + \frac{15}{100}) * (1 + \frac{-15}{100})\right)^{\frac{1}{4}} - 1$$

$$z \approx 6,6\%$$

* * *

The Essentials

The **absolute change** of a variable X is : $\Delta X = X_{final} - X_{initial}$

The **relative change** of a variable X is :

$$\frac{\Delta X}{X} = \frac{X_{final} - X_{initial}}{X_{initial}}$$

The relative change of a variable X stated in % is :

$$\frac{X_{final} - X_{initial}}{X_{initial}} * 100$$

The **multiplying factor** of a variable X which has increased by a% is : $(1 + \frac{a}{100})$
so that $X_{final} = X_{initial} * (1 + \frac{a}{100})$

The **aggregate change** of two successive increases (decreases) by a% and by b% is :

$$\left(1 + \frac{a}{100}\right) * \left(1 + \frac{b}{100}\right)$$

So that :

$$X_{final} = X_{initial} * \left(1 + \frac{a}{100}\right) * \left(1 + \frac{b}{100}\right)$$

The **average growth rate**, noted $c\%$, of two successive increases (decreases) by $a\%$ and by $b\%$ is

$$\left(1 + \frac{c}{100}\right) = \sqrt{\left(1 + \frac{a}{100}\right) * \left(1 + \frac{b}{100}\right)}$$

$$c = 100 * \left(\sqrt{\left(1 + \frac{a}{100}\right) * \left(1 + \frac{b}{100}\right)} - 1\right)$$