www.alea.pt

Educational Dossiers

Notes on the History of Statistics

Maria João Ferreira
Isabel Tavares
With contributions from Professor Maria Antónia Turkman

1. Foreword

The ALEA project – Local Applied Statistics Initiative – contributes toward the creation of new statistics teaching support media for primary and secondary students and teachers.

The project arose from an idea jointly fostered by Escola Secundária Tomaz Pelayo secondary school and Instituto Nacional de Estatística [National Statistics Institute of Portugal], founded on the requirements and structures that the intervening parties possess. The improvement of statistical literacy is thus a significant proviso in guaranteeing the provision of a service of public value. The teaching of statistics in lower and upper secondary education constitutes one of the most important instruments aimed at achieving this objective. The address of ALEA's web page site is www.alea.pt.

Contents

The Educational Dossiers area was designed to support the creation of educational material on a range of topics.

Dossiers available in the English version of ALEA:

- Notes on the History of Statistics
- The Cinema in Numbers
- Graphical Representations
- Statistical Surveys

The **Notes on the History of Statistics** dossier now follows. This dossier covers some of the most important moments in statistics and its history, from the ancient civilisations to the present day. Some scientists, though not all, that made a significant contribution to the development of statistics are also mentioned in this dossier. The final chapter contains an overview of the development of statistics in Portugal. The ***See Also*** section at the end of the dossier contains links to other studies of interest related to the topics covered (publications and web sites).

2. Ancient Civilisations

2.1 Introduction

Statistics has existed since the start of civilisation - from the obscurest of ancient times up to the more visible modern world.

Counting, calculating and carrying out censuses have always been a constant concern of all cultures. In the civilisations of ancient Greece, Rome, Egypt, Israel, India, Japan and China, as well as others, the State needed to know the size of its population, for both economic and social reasons. The emperors of the time ordered censuses of the population, with the purpose of collecting taxes and recruiting military personnel, since wars were constantly occurring and young men had to be found to be physically trained for war.

Those who did not participate in censuses in ancient civilisations were condemned to death.

These censuses cannot be compared to those carried out nowadays since they were not based on statistical principles nor were they exhaustively implemented. Nevertheless, we can say that statistics started with these societies, not in the form that we currently know it, but in a simpler and more elementary manner.

2.2 Ancient Civilisations and Censuses

2.2.1 Ancient Greece (2100 BC to 146 BC)

Ancient **Greece** covered a vast area. It was formed of a set of politically autonomous city-states, possessing a common language and customs. In the 5th century BC **Athens** stood apart from all the other city-states. It possessed the most accomplished culture of all the Greek cities, in particular in the arts, theatre, history and philosophy. It also possessed the most democratic government of all the Greek cities. Sparta and Corinthia were other cities of note besides Athens.

As Bedarida et al. (1987), stated, Athens was the Greek city that most accurately knew its population. Aristotle informs us that a measure of top-quality wheat (wheat producing snow-white flour) was offered to Athens' priestess on the birth of each child and a measure of barley was offered on each person's death. In addition, every young man was registered as a citizen on turning 18 years of age and placed on the list of men ready to bear arms. Until they were 18, they just studied arithmetic, literature, music, writing and physical education. Young women did not receive any kind of official education, instead they learnt the domestic chores and manual tasks with their mothers. These descriptions provided by ancient historians provide us with evidence of the first censuses carried out by ancient civilisations. We also know that foreigners were also included in censuses, by means of the specific taxes they had to pay, which was levied per person.

It is curious to note that in descriptions of Athens, Aristotle would not only describe the specific situation of a city's or country's government, justice, sciences, arts, museums and customs, but would also compare these areas with those of other states. Thus, in this part of Aristotle's work we have the beginnings of **Descriptive Statistics**.

Definition of Descriptive Statistics:
The descriptive study of the data of a sample (or of a population) where all the information collected is summarised by means of tables and plots, calculating some of the data's characteristics, such as the mode and mean, for example.

2.2.2. Ancient Egypt (5000 BC to 30 BC)

<u>**Location**</u>

Ancient **Egypt** occupied almost the same area occupied by present day Egypt. Its civilisation, which was very close to the River Nile, was almost completely surrounded by desert.

[Key:
Mediterranean Sea, Lebanon, Syria, Israel, Jordan, Alexandria, Suez Canal, Sinai Peninsula, Saudi Arabia, Libya, Egypt, Nile, Red Sea, Lake Nasser, Sudan]

Egyptian culture is one of the oldest and most enduring, having lasted for almost five millennia. It benefited from an abundance of good land, nearby mineral resources and a good strategic position.

2.2.2.1 Censuses and "mass" statistics

If calculation goes all the way back to the first human communities, then “mass” statistics started with the great empires of the ancient world, which were concerned with managing their assets, men, weapons and their colossal public works. Such computation assumes complex organisation and a sound administrative structure. Censuses were already implemented by one of the most ancient civilisations known - Egypt, caused in part by the lack of labour required for the construction of the pyramids. A record on the Palermo Stone dated from 2900 BC does, in fact, refer to the undertaking of a census of the population. Between 2700 and 2500 BC biannual censuses, which were later performed yearly, were being performed relative to the different assets liable for taxation. Around 1900 BC, lists of the members of soldiers’ families started to be compiled, for tax and military purposes. Lists of houses, of the heads of families and their relatives, indicating the name of the father and mother of each occupant of the house, began to be compiled in the middle of the 13th century BC. During the reign of Amasis II (6th century BC) all individuals had to declare their profession and sources of income every year to the government of their province (failure to do so would warrant the death penalty).

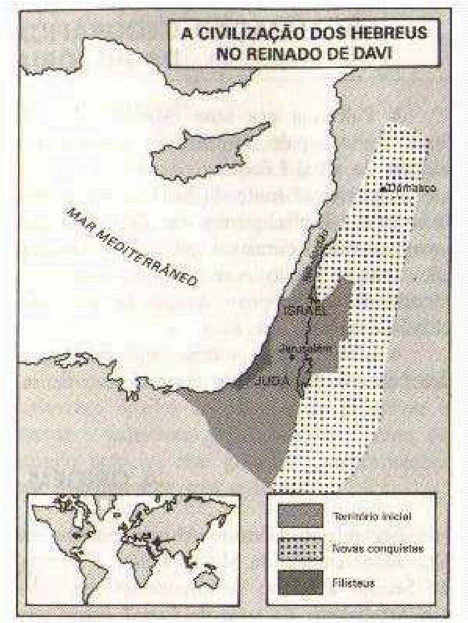
Census: The scientific study of a population of people, institutions or physical objects in order to obtain quantitative knowledge regarding important characteristics of that population.

The ancient Egyptians believed that they could communicate with the gods through their king. The king held absolute power, running the government, trade and foreign policy, and passing laws and leading the army.

All workers paid taxes, which were calculated as a percentage of their production. In addition, each house had to make one person available for a few weeks each year to work on public works. The pyramids were probably constructed by workers providing such annual service. Moreover, the quality of the construction and sheer size of the pyramids implied the organisation of human resources never before demonstrated in any other civilisation.

The administration of the Egyptian state was only possible thanks to a very large and highly effective civil service, primarily composed of numerous scribes. These scribes used hieroglyphic characters that had appeared in Phoenicia in 3000 BC. Hieroglyphs were based on pictures and were used right through to the end of the Egyptian empire.

2.2.3. Israel and the Book of Numbers (1700 BC to 70 AD)



People mix up the terms *"Hebrew"*, *"Jew"* and ***"Israelite"***. The Hebrews were the first Jews, the first inhabitants of the Land of Israel, those that first used the Hebrew language. The term has a meaning that is more ethnic and tribal than religious.
A distinction was made between Israelites and Jews in the period between the 10th and 8th century BC, when 10 tribes established themselves in the north of the Holy Land (the Kingdom of Israel) and two in the south (the kingdom of Judea). These terms are, however, synonymous nowadays.

[Key: The Hebrew Civilisation in the Reign of David;

Mediterranean Sea; Damascus; River Jordan; Israel; Jerusalem; Judea.

Initial territory; New Conquests; Philistines.]

"Pour une Histoire de la Statistique" (Bedarida et al., 1987), states that the attitude of the Hebrews to censuses was largely responsible for shaping Western opinion for almost 2000 years.

The Hebrew cultural legacy was important in the formation of various traits of Western culture, since Hebraic cultural production is connected to religious life.

We have also absorbed Hebrew culture and inherited their belief in one God, creator of the entire Universe and all that it contains. They are responsible for writing a large part of the Bible. Thus, the history of the Hebrew people cannot be dissociated from the history of their religion, even more so since what we know about the Hebrew people was mainly taken from the Bible, particularly the Old Testament. This being so, the referred-to work highlights the curious fact regarding the ambiguous, hesitant and contradictory attitude verified in the Bible. Censuses were mainly deemed to be sacrilegious since they were seen to be intervening in the secret of life and creation, which was the sole domain of God. It is evident that here, as in other places, the population feared being subject to a census for military and tax purposes, and, on the other hand, it seemed to them that making inventories of their wealth, both in terms of men and goods, could lead to ruin.

Therefore censuses, for all these reasons, did not seem to be permitted unless they had been ordered by God himself. They were also sometimes attributed to Satan, which seems to be the only way to explain the misfortunes that could happen, such as the lives of people taking part in a census could not be saved and were forever condemned.

Censuses do not seem to be permitted, unless ordered by God. In addition to this, they were often attributed to Satan, as this seemed to be the only way to justify the misfortune that would occur, where the lives of those participating in a census could not be saved and they were condemned forever.

The census ordered by God at Sinai is cited in two passages of the Book of Moses, which has the name the "Book of Numbers". Yahweh ordered Moses in the desert at Sinai "to take a general census of all the congregation of the children of Israel, clan by clan, family by family" (Numbers, 1, 2). In the Book of Exodus (30, 12-15), it is written that when Moses took the census of those that should be counted, "each one must pay Yahweh for the salvation of their life, so that this census might not bring disaster on them". Yahweh demanded that homage be paid and offerings made exclusively to Him and, in return, He would be the all-powerful protector of the Hebrew people.

2.2.4. The Chinese Census Machine

Location

China is located in the extreme south of the Asian continent. The country is crossed by a number of great rivers: the Yellow and Blue rivers, along with other rivers, such as the White and Red rivers, form long valleys that fertilise the fields in the heart of China.

Ancient civilisations that were wholly accustomed to taking censuses also developed in the Far East.

The oldest historical records that we have tell us that the first census was undertaken in 2238 BC by the first emperor of China, Yu or Yao. The Chinese court wanted to know the exact number of inhabitants in the kingdom so that the land could be divided up and distributed and tax registers drawn up and military recruitment undertaken.

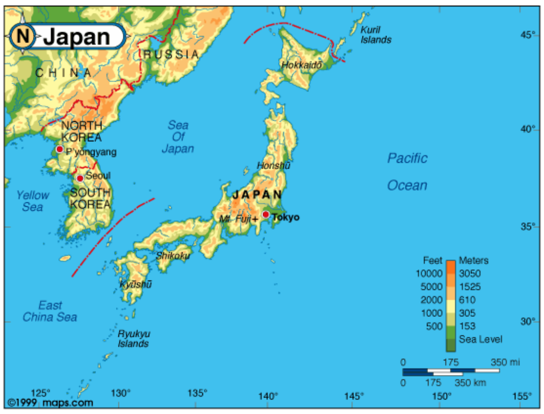
A number of censuses were performed in China.

1. <u>Censuses linked to a recruitment initiative</u> (during the era of the Han dynasty, 200 BC – 200 AD). The State, as a centralised body, sought to evaluate the number of soldiers available for wars and public works.
2. <u>Censuses linked to the land distribution system</u> (from the third kingdom to the fifth dynasty: 221-959 AD). In order to encourage agricultural production and restrict large-scale domains, the central government effectively redistributed land in exchange for services and as fixed-term payment, thereby bringing to the fore the need to know the size and make-up of families.
3. From 960 to 1368 AD <u>the main objective of censuses was to collect taxes</u>. The notion of family still prevails.

   During the Ming Era (1368-1844 AD), the “admirable census machine”, so described by M. Cartier, was in operation. “<u>Record charts</u>” of the population were compiled right up to the end of the dynasty. These records contained the name, profession, gender and age of each person.
4. From 1644 AD onwards (Ching) there was period of <u>registration with the police, in order to watch over the movement of people and track down undesirable individuals</u>. In 1741 the estimation methods are modified. In 1975, the *pao-chia* system was in force, which required that all houses display a poster specifying the number of occupants, gender, age, profession and the amount of tax paid. This system allowed demographic series from 1750 to 1850 to be obtained.

In conclusion, for a long period of time the immense Chinese empire made every effort to perform censuses, despite the difficulties, with “patience” comparable to the scientific meticulousness of modern countries.

2.2.5. Japan to the Tokugawa Era



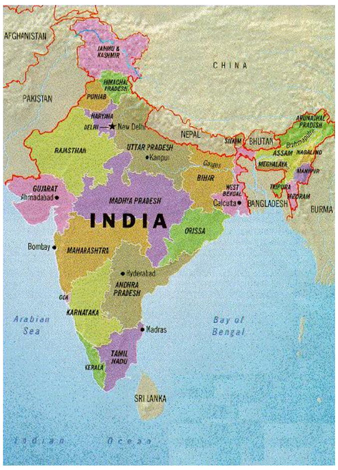
<u>Location</u>

Japan is located on the eastern edge of Asia. It is made up of four main islands and three thousand smaller islands.
The country is subject to earthquakes and volcanic eruptions. It possesses the world’s second largest economy.

It appears that censuses were carried out in Japan in ancient times, but the results of these censuses were not published. The first census took place in 86 BC, during the reign of Emperor Soujin. The economic activity of the population at the time of the census was recorded, in order to examine its evolution. In the middle of the 7th century BC, Taika's reforms, which were aimed at making the entire population liable to pay a tribute, coincided with land re-distribution, which required the establishment of land records and the registration of civil rights, reviewed every six years. The data on families was collected by the local council and it was archived according to the family's resources, with further sub-divisions according to gender and age group. The purpose of this census was not only to levy taxes but also to facilitate recruitment for the army and for forced labour.

According to written sources, local censuses were performed at the end of the 17th century (1665), during the reign of the Tokugawa (17th to 19th century). The first general census was performed in 1721, and then repeated every six years afterwards. Specific categories of the population were excluded from this census - the nobility, the poorest people and children under the age of 15. This record, as is evident, contained a number of gaps. Based on this, the urgent Japanese need to undergo demographic development is comprehensible.

2.2.6. A Census Treatise in Ancient India



<u>**Location**</u>

India is a diamond-shaped country located in south Asia. It is bordered to the north by China, Nepal and Bhutan, to the east by Bangladesh, to the north-west by Pakistan and to the south-east, south and south-west by the Indian Ocean.

Another well known example of the interest expressed by ancient Asian empires in counting their population is the treatise drafted by the Hindu Kautilya, a minister of King Kandragupta

(313-289 BC), founder of the dynasty and the first Indian empire - Maurya (313-226 BC), in the 4th century BC. This treatise was extremely original and advanced for its time. It also a treatise on economics, as it encompasses political sciences; its correct name is *Arthasastra*, in other words the treatise or science (*sastra*) of progress (*artha*).

The work describes the centralising and expansionist state that was the Maurya empire, and Kautilya (later Machiavelli) reflects on the art of governing and indicates to the sovereign how to incessantly increase the kingdom. According to Kautilya, the state should manage and control everything. The state must be the absolute master of the economy, governing with the aid of a very extensive administrative apparatus, performed through the army and secret police. In order to fulfil a "planning role", the state, according to Kautilya, must rely on censuses, statistics and registers. "Everything that is done must be known by the state: from the active population to the number of elephants, including raw materials, manufactured products, prices and wages".

Arthasastra: The Treatise of Progress

In Arthasastra, Kautilya describes with great precision the duties of government superintendents at the different hierarchical levels throughout the kingdom. In each state, the superintendent must divide the country into four provinces, perform a census and write down the number of villages and rank them according to wealth (rich, average and poor) so that the labour and products which, for the most part, were supplied in the form of taxes, could be more easily accounted for. These guidelines, on the other hand, also aided the task of recruiting of soldiers.

The provincial superintendent ensured that written records were kept, in particular with regard to households and those paying taxes. The names of the people belonging to each one of the four classes (*varsa*), the number of manufacturers, shepherds, traders, artisans, free labourers or slaves, the number of animals as well as the amounts of money, work, rights and fines were also recorded. The superintendent also recorded the number of men and women, children and aged persons in each family as well as their professions, lifestyle, income and expenditure.

The general governor of the country registered the number of inhabitants, their gender, caste, family names and professions, as well as the place of residence, income and expenditure.

The state, according to Kautilya, which was informed and supplied with such details, would then be able to more effectively perform its role of forecasting and economising.

2.2.7. Censuses in Rome (750 BC to 476 AD)

Location

The city of Rome was the cradle of Roman civilization. Is is now the capital of the Italian Republic, which extends out into the centre of the Mediterranean Sea. Italy has two large islands, Sicily and Sardinia, to its south and west.
Around 80% of the country is mountainous or hilly, with the largest expanse of flat land located on the Padania plain, crossed by the River Po.

The concepts and practices concerning censuses in the city of Rome were strongly influenced by Eastern thinking. At the end of the 6th century BC censuses were performed every five years until the year 68 BC when, after an interruption of 20 years, they recommenced under the reign of Augustus, being performed every decade.

According to tradition, the first census authorised the division between civilian and military tasks not by head, but according to wealth.

Roman citizens were obliged to declare their wealth, their name, the name of their parents, their age, the names of their wife and children, the tribe where they lived and the number of slaves they owned. If they did not provide all of this information they could be deprived of their assets or their rights as a citizen.

Censuses not only allowed citizens to be classified according to their income but they also permitted taxes to be levied on assets and the social status of citizens to be identified, which would determine citizens' appropriateness for political and military posts in cities.

2.2.8. Statistics at the Time of Christ…

The date of birth of Christ is, nowadays, the subject of a fair amount of controversy, given that the Roman governor of Syria, which included Judea and Galilee, who was ordered by the Senate to perform a census, used what is probably the most absurd method of all to do this (Collected Works of J. Tiago de Oliveira, Volume II, 1995). The Bible states that St Joseph and the Virgin Mary had to leave Nazareth in Galilee and travel to Bethlehem in Judea in order to be included in the census that had been ordered by Caesar Augustus (the people had to be interviewed in the town of their birth). Jesus was born while they were at Bethlehem.

3. ... Up to the Modern Age

3.1. Statistics and Games of Chance

Games have always greatly interested people and they have been widespread in all civilisations. They were so important that in the Greek pantheon of gods there was a goddess of the arts of chance, Thyke, the equivalent of the goddess of fortune of the Roman pantheon of gods, who was known by all through her symbol, the wheel of fortune (Oliveira, 1995). The Portuguese term for 'chance', or more correctly 'luck', in this context (*azar*) does not mean 'bad luck' or 'ill fortune'; *azar* comes from Arabic and means 'chance'.

The term for "chance", used in the Portuguese expression for "games of chance" does not mean bad luck or ill fortune but simply means chance.

In Portugal it is written down in *Diario da República* (the Official Gazette of the Portuguese Government), and therefore constitutes law, that that the Portuguese term for games of chance (*jogos de azar*) simply means this and not that they are games of bad luck

3.2. The Start of Probabilities

J. Tiago de Oliveira states that in Jerusalem there is still a grid drawn on the ground of the prison in which Christ was kept, forming a square divided into nine equal parts, used to play the old game of noughts and crosses. Games have been ever present in nearly all civilisations, as evidenced by many archaeological or historical documents. Curiously, games were never studied until the Middle Ages.

The mathematical tackling of luck, ill fortune and risk only began less than 500 years ago. The discipline that was created, **Probability Theory**, was born from the attempts made to quantify **insurance risks** and to assess **the possibilities of winning games of chance**.

As the Middle Ages drew to a close, the growth of urban centres lead to the appearance of life insurance. The first mathematical studies of insurance were carried out because of these occurrences. But insurance mathematics only reached a sufficiently mature state after 250 years of study, thanks to the work of the mathematician **Daniel Bernoulli**. He tackled the classic problem of calculating the expected number of survivors after *n* years of a specified number of

newborn babies. **Bernoulli** also took the first steps towards new types of insurance as he calculated the death rate caused by smallpox in people with a given age.

Girolano Cardano (1501-1576) was a notable mathematician, notable swindler, notable doctor and notable scholar in the field of algebra and probability. He wrote a small manual on games of chance, *Liber de Ludo Aleae*, which is probably the first book on probability that analyses games and odds. Cardano was the first person to use **combinatorial techniques** to calculate the quantity of favourable possibilities in a random event. He limited himself to solving a few concrete problems, in other words, problems with strictly numerical data, but he never produced any theorem.We can consider **Pascal** (1623-1662) and **Fermat** (1601-1665) to be the founders of **Probability Calculation**.

Blaise Pascal was born in Clermont in 1623.A philosopher, mathematician, physicist, theologian and writer, he greatly contributed to the study of probabilities, discovering new properties of the arithmetic triangle, which is nowadays known as **Pascal's Triangle**.

Combinatorial techniques: A counting technique that indicates the number of possible results of an experiment. The actual results are of no interest (direct counting), just the total number of results.

The first major probability problem, which was posed to **Pascal** by the **Chevalier de Méré**, arose in the court of the King of France, where one of the nobility's amusements was gambling. The problem concerned a specific game with three dice, which **Méré** was unable to understand the results empirically observed. **Pascal** and **Fermat**, separately found the solution to the problem, but **Pascal**'s solution was very specific, while **Fermat**'s was probably the first ever general method to determine probabilities.That problem produced two situations with the same probability, but which differed in the empirical verification of the frequency analysis.This is where the idea of the **Law of Large Numbers** starts to emerge, along with the 'automatic' identification of probability and of frequency in a large number of tests.

Pierre de Fermat, was born in Beaumont in 1601. He was known as the "Prince of Mathematics' Amateurs", studying mathematics as a pastime. He was a lawyer by profession, as well as an advisor to the Toulouse Parliament from 1631. He is considered to be the creator of the theory of numbers and the forefather of analytical geometry, the calculation of probabilities and differential calculus. His contribution to the field of probability calculation derives from the correspondence established with his colleague Pascal as they tried to solve the problems posed by the Chevalier de Méré.

This is the start of a period, which terminated at the start of the 20th century, in which statistics takes a back seat to the development of **Probability Calculation**.

In simple language, the **Law of Large Numbers** tells us that the frequency of an occurrence in a long series of tests moves increasingly closer to the probability of that occurrence, a probability that thus arises as a limiting frequency. In other words, the Law of Large Numbers is expressed through the idea that if the probability of a number being thrown with a six-faced die is 1/6, in 100 successive independent throws this number will appear 100/6 times, in 1000 successive independent throws this number will appear 1000/6 times, and so on.

3.2.1 The Oddity of "Passe-dix"

"Uncertainty has been a concern of Man for a very long time. And it was the pastime of games that, through probability, constructed the instruments and rules that allow statistics to measure the level of uncertainty (or implementation) of phenomena" (Oliveira, 1995).

"Passe - Dix"

Passe-dix was a common game in the French Court. It is a game where the player throws three dice simultaneously and wins if the sum of the three dice is greater than ten and loses if it is nine or less. At the time of Louis XIV, the Chevalier de Méré, who was an intelligent and chronic gambler, observed that "11" was thrown much more than "12", which seemed strange to him since the ways of obtaining a total of 11 and 12 are as follows:

Table 1

Total of 11	Total of 12
6+4+1	6+5+1
6+3+2	6+4+2
5+5+1	6+3+3
5+4+2	5+5+2
5+3+3	5+4+3
4+4+3	4+4+4

and they are equal in number (6), which should produce an equal or very similar frequency. Nevertheless, it is easy to see that while the combination (6,4,1) can be produced in six different ways (this is easier to understand if, for example, each die has a different colour, then 6,4,1 can be produced by a "6" with the white die, "4" with the blue die and "1" with the green die, or by

a "6" with the blue die, "4" with the green die and "1" with the white die, etc. - six different ways in total), the same is not true for the combination of (4,4,3) which can only be produced in three different ways, where the "3" is generated by one of the three dice and the "4" by the other two. If the calculation is performed in this manner (with the total number of different ways of forming each combination in brackets after each one) we can observe that "12" can be generated in 25 different ways while "11" can be generated in 27 different ways. Thus, de Méré had correctly verified that in the game of "passe–dix" a total of 11 was more frequent (more probable) than a total of 12, contradicting that which appears to be the case when the problem is first looked at.

Table 2

Total of 11		Total of 12	
6+4+1	(6)	6+5+1	(6)
6+3+2	(6)	6+4+2	(6)
5+5+1	(3)	6+3+3	(3)
5+4+2	(6)	5+5+2	(3)
5+3+3	(3)	5+4+3	(6)
4+4+3	(3)	4+4+4	(1)
	(27)		(25)

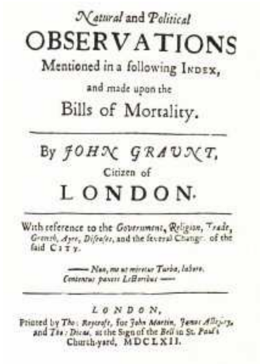
3.3 The Development of Statistics

Statistics started moving towards the science that we know today in the 18th century.

Two schools appeared at that time, one in Germany and the other in England. The German Descriptive School, as it became known, moved away from the ideas that provide the basis for modern statistics. The most well known representative of the German School was **Gottfried Achenwall** (1719-1772), considered by some authors to be the father of the word "statistics". But, in the opinion of Sir Maurice Kendall (Pearson and Kendall, 1820), this word had already been used in Italy in a work by the historian Girolamo Ghilini, in 1589, referring to a register of "*civile, politica, statistica e militare scienza*" [civil, political, statistical and military science]. According to Kendall, the word used in the German School only referred to the method used in studies aimed at describing political states and, if any numerical information appeared in these records it was purely coincidental. The English school, the "School of Political Arithmetic", specialised in the numerical study of social and political phenomena.

The School of Political Arithmetic was concerned with the numerical study of social and political phenomena, while the German School only described states.

The English school produced two statisticians that were important in the development of **modern statistics**: **John Graunt** (1620-1674) and **William Petty** (1623-1687).


Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.
By JOHN GRAUNT,
Citizen of
LONDON.
With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Changes of the said CITY.
—Non, me ut miretur Turba, laboro, Contentus paucis Lectoribus —
LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

Front cover of John Graunt's published work

The work of **John Graunt** (Seneta and Heyde, 2001) constitutes the basis of modern statistics. **Graunt** studied the mortality rate in the city of London and the prevalence of natural, social and political causes of this phenomenon. Graunt carried out an exhaustive analysis of the number of people dying from a range of illnesses and he estimated the number of births of males and females. He was the first person to make observations on gender and he showed that more men than women were born and that for every 100 people born, 36 would die by the age of six and seven would survive to the age of 70.

Graunt published his work *Natural and Political Observations Made upon the Bills of Mortality* in 1662. This work greatly advanced the quantitative analysis of social phenomena and the development of Demographical Statistics.**John Graunt**'s work attracted the attention of King Charles III of England, and he proposed **Graunt** as a founder member of the Royal Society. **William Petty** worked together with **John Graunt** for three years and he too can be deemed a motor of modern statistics.

John Graunt was born in London in 1620. He was a respected and studious man, and held many important posts in the city of London. He inherited his father's shop and managed to turn the business into a great success. He was a Captain of the military band and then a Major. He was one of the founders of the Royal Society, and lived in an era noted for the birth of modern science. In 1662, Graunt published his magnum opus *Natural and Political Observations Made upon the London Bills of Mortality*, which was the first time demographical data had been statistically analysed and the first attempt at applying the theory to actual problems.

Before the General Register Company was created in England, **Petty** had already proposed the creation of a central statistics company. The purpose of this company was not only to register births, deaths and marriages, but also house types, family sizes, and the gender, age, profession and level of education of each member of a family. It proposed the creation of Survival Bills, based on the death rates of each age group. The connection of probability to statistical knowledge gave a new dimension to statistics. A new phase is deemed to have started, in which **Statistical Inference** commences. A number of scholars come to the fore during this period. These include **Christian Huygens** (1629-1695), who introduced the notion of **mean value** or **mathematical expectation** in 1654. Another was **Abraham De Moivre** (1667-1754), who performed the groundwork for the development of analytical geometry and probability theory on publishing the celebrated *Doctrine of Chances* in 1718, about the theory of chance. This work contained a definition of statistical independence as well as many problems related to dice and other games, including the probability of selecting different coloured balls from an urn, for example. **De Moivre** is credited with the principle that states that the probability of a compound occurrence is the product of the probability of the components, even though this idea had already appeared in previous works. He was also interested in demographic statistics and he created the theory behind pensions.

> **Statistical Inference**: A fundamental phase of statistical analysis, during which, once certain properties are known (obtained via a descriptive analysis of the sample), expressed by means of propositions, more general statements are formulated, which express the existence of laws (relative to the population).

But the three great figures of Probability Theory are, without a doubt, Jacob **Bernoulli**, Thomas **Bayes** and Pierre Simon **de Laplace**.

Jacob Bernoulli (1654-1705), who published *Ars Conjectandi* after his death, in 1713, showed, at the same time as **Leibniz**, a glimpse of what statistical science was going to be or should be. One of the great contributions to statistics was the **Bernoulli distribution**, which states that each attempt has two possibilities of occurrence, called success and failure (e.g., a coin thrown into the air will land heads or tails). This distribution was the basis of the binomial distribution.

> **Bernoulli Proof**:
> 1. A fixed number of observations, n, is considered, which are usually called trials;
> 2. The observations are independent from one another;
> 3. Each observation can produce one of two possible outcomes, which we call success or failure;
> 4. The probability of success, ***p***, is constant from observation to observation.

All of these contributions were extremely important to statistics since they started to shed light on Probability Theory's great problems. These problems were only fully, methodically and systematically solved in 1933 by **Kolmogorov**.

Bayes (1701-1761) came to the fore later on. He was, according to Tiago de Oliveira, the first to clearly set down the fundamental problem of statistics: how it is possible, based on

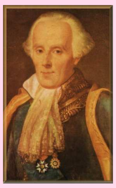
observations, to know something with regard to a specific universe. In 1762 **Bayes** demonstrated the method that came to be known as the **Bayes Rule**, which consisted of dividing the sample set into different subsets of known probabilities. The rule is represented by the following formula:

$$P(A_i/B)=\frac{P(B/A_i)P(A_i)}{\sum P(B/A_j)P(A_j)}$$

The ideas of **Thomas Bayes** were not well accepted by scientists of the time given that the equations resulting from Bayesian statistics were often very difficult to solve. Nonetheless, from the 1990s onwards these ideas were given new life with the growing use of computers, and they are often used in statistical studies nowadays.

In the meantime, another figure of great importance, Pierre Simon de **Laplace** (1749-1827) published *Théorie Analytique des Probabilités* [Analytical Theory of Probability] in 1812, which is deemed to be a great milestone in probability theory. **Laplace**, in this work, defined probability as being the number of times that a given occurrence can occur, divided by the total number of cases that can occur, where these have an equal chance of occurring.

"...It is remarkable that this science, which commenced with studies on games of chance, has attained the highest levels of human knowledge".
-- Laplace



Pierre Simon de Laplace was born in Normandy (France) in 1749. A French astronomer and mathematician, he studied at Beumont-en-Auge, where his interest for mathematics was awoken. His great contribution to the development of statistics was the publication of the treatise, Analytical Theory of Probability, in which he described a useful calculation to guarantee a "degree of rational credibility" regarding propositions of random events.

Tiago de Oliveira (1995) states that in more backward countries statistics is often reduced to the accountancy of facts and the listing of events, such as the number of people that died from illness A or B, without any analysis of the cause of these facts.

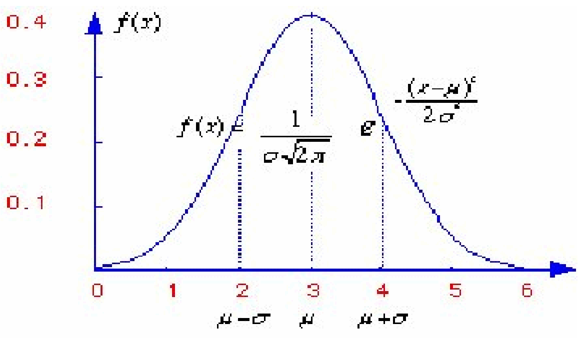
"...all observational sciences went through the same phases in the beginning; they were arts because they just grouped together collections of facts belonging to the same order of things, in a more or less convenient fashion. Through the comparison and study of those facts they were subsequently elevated to the position in which we see them shine today. Why be more demanding with statistics?"
-- Adolph Quételet

The first person to tackle the problem of the lack of organisation in the collection of data and to propose the creation of an independent statistics service was the Belgian **Adolph Quételet**

(1796-1874), who proposed the organisation of censuses and performed the organisational groundwork of the Belgian statistical service. **Quételet**'s work encouraged the more widespread use of the normal distribution beyond its applications for analysing errors, in particular, its use in the study of human characteristics, such as height and weight.**Quételet** improved data collection methods and he worked on the statistical analysis of data related to crime, death, geophysics and astronomy. He organised the first statistics conference in 1853 and wrote *Sur l'homme et le développement de ses facultés, essai d'une physique sociale*, which was published in the same year.

Another mathematician that greatly contributed to the development of statistics was Carl Friedrich **Gauss** (1777-1855), known as the "Prince of Mathematicians"**.** He provided the starting point for some of the main research areas of modern mathematics; he formulates **Gauss's Law**, which concerns the distribution of certain values along a bell-shaped curve. This was an extremely valuable contribution to the field of statistics.

Example of a Gauss Curve



The normal distribution is an approximation to the distribution of the values of a characteristic. The exact form of the distribution depends on the mean and standard deviation of the distribution.

Two equally important figures in the development of statistics were: Siméon Denis **Poisson** (1781-1840), who discovered the limited form of the binomial distribution in 1810, which was later named after him; and **Marquis of Condorcet** (1743-1794) was the first person to apply the "magical arts of chance" to social problems and methodically analyse the problems of voting. These two men were the first to concern themselves with applying statistics to social fields.

An explosion in the development of modern statistics took place starting in the second decade of the 19th century, principally due to **Ronald A. Fisher**, who is now known as the father of modern statistics.

We will find out more about this famous mathematician in the next chapter.

4. Modern Day Statistics

4.1. Introduction

The turning point of descriptive or graphical statistics to the methodological study occurred in the second half of the 19th century. This process commenced at the first international statistics conference held in Brussels in 1853 (Oliveira, 1995). Statistics was considered, up to that point in time, solely for use to compile data, arrange it in tables and calculate means and perform other simple statistical calculations. It was not deemed to be anything more than this. Statistical decision making was, more often than not, performed intuitively, and the subsequent value calculated from the sample served only to see if it was close or far away from the value that was theoretically expected. New names that were significant in the development of statistics emerged at this time: **Galton**, **Karl Pearson**, "**Student**", **Lexis** and **Von Bortkiewicz**. These mathematicians "opened" the way for **Fisher**, **Neyman** and **Wald** to set down the basics of modern statistics, to search for optimal inference methods, to study inductive behaviour and to make inductive and fuzzy comparison more meticulous.

4.2. Statistics in the Study of Human Hereditariness

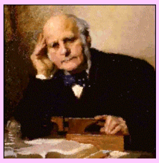
The 'fathers' of statistical inference in the field of hereditariness were J. **Neyman** and Karl **Pearson**. Even though their studies were associated to questions related to the biology and genetics fields, the methods that they created, such as the null hypothesis and significance level, have become a routine part of the work of any modern statistician and scientist using statistics.

Francis Galton, one of the great founders of modern science and human science, in particular in the 19th century, was the founder of anthropology, which encompassed the study of humankind and its origins. He also produced many meteorology studies (he discovered and invented the term 'anticyclone') and he provided the bases for the commencement of the study of genetics.

He created the term 'eugenics' and was actively involved in its practice. Eugenics studied the genetic improvement of the human species and **Galton** believed that the physical and mental characteristics of human beings were due to hereditariness. He designed instruments to measure sense capacity, memory and imagination. In 1865 he published the book *Hereditary Talent and*

Eugenics was the term defined by Francis Galton to encompass the study of agents under social control that can improve or weaken the physical or mental qualities of races in future generations.

Genius, in which he proposes the idea that intelligence is predominantly inherited and not the result of environmental influence. **Galton's** main objective was to prove how character and talent are transmitted by reproduction through successive generations. He set up his laboratory in London and visitors could be examined by parading in front of his instruments. He measured height, weight, the width of the palm, respiratory capacity, strength, and many other parameters. He used the data collected to draw up tables, probability curves, mean values and perform other calculations. **Galton** created an explanatory diagram, which would later on be replaced by the measure of **correlation between two variables**. **Pearson** would be the scholar who, later on, would create the **correlation coefficient**. Around 1870, **Galton** decided to modify a device that he had created and used in lessons to illustrate the basics of the law of errors. He called this device a *quincunx*. (See box for description.)



Francis Galton was born on 16 February 1822 close to Birmingham, England. It is recorded that before he was three years old he was able to read a simple book. He provided proof of his genius for mechanics and mathematics from a very early age. He founded the biometry school and became interested in statistical methods and their application to all fields. Galton's works are based on quantitative measures originating from Gauss's normal law. His main contribution to statistics is the concept of correlation and its measurement through the **correlation coefficient**.

Galton modified the *quincunx* to demonstrate that normal distributions were usually a **mixture of normal distributions**. In other words, by means of experimentation and the device he had invented, called a *quincunx*, he concluded that he had clear experimental proof that the significant causes of the phenomena could in fact be isolated according to the law of error.

Initially **Galton** used the natural world for inspiration, focusing on fruit orchards, and how specific factors, such as appearance, can affect fruit size.

Galton's Quincunx[1]

This device consists of a set of lead balls that descend a steep ramp. During their descent they collide with pegs placed all along the ramp.

It is not difficult to imagine conditions in which the probability of the balls rebounding to the left or the right of the peg is equal. If two pegs are placed downhill of each peg, horizontal to each other, and the slope of the ramp is correctly adjusted, the ball will hit one or the other of these pegs after rebounding off the first peg. Once again the probability of the ball rebounding to the left or the right of these pegs is equal.

The probability of rebounding to the left off both pegs or between them or to the right of both is in a ratio of 1:2:1. The process can be continued and it is clear that the probabilities of a ball passing between the different pegs of a row are proportional to the numbers of Pascal's Triangle:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
… … … … … …

The distribution of probabilities along the *n*th row is thus proportional to the coefficients of $(1+t)^n$. Such a distribution is called a binomial distribution.

A ramp of this kind is called *Galton's Quincunx*, in homage of its inventor, Galton; 'Quincunx' is the Latin name for the number five of a die, or for any similar standard.

Slots for the balls were made at the bottom of the ramp and a piece of glass was used to separate the slots and ensure balls do not pass from one to the other. The upper part of the ramp has a storage area to hold the balls. This is closed off by a small door that can be removed. When the door is removed, the balls roll down the ramp and are diverted by the pegs set out in a suitable manner along the ramp. If the angle is correct, the number of balls in the respective slots is very close to the binomial distribution.

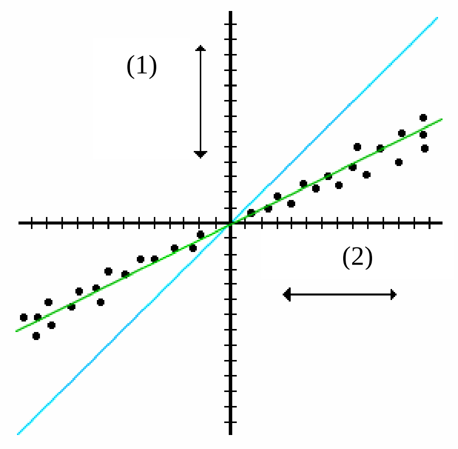
With a large number of balls and many rows of pegs, the result is similar to the standard error curve $y = ke^{-\frac{x^2}{2s^2}}$, where *k* and *s* are constants. The curve formed by the columns of balls in the slots should provide a rough idea of its shape.

4.2.1 Law of Regression towards the Mean

The British researcher, Francis **Galton**, put forward the law of regression towards the mean based on his work with parent-children pairs.

[1] Natália Bebiano, 1999

Galton's Law of Regression towards the Mean:



[*Key*: (1) Children's values; (2) Parent's values]

The graph above indicates the relationship of a metric variable between parents and children (height, for example). The blue line represents that expected if the children had exactly the same mean value as the parents. Note that parents with greater values of the characteristic have descendents with a mean value of that characteristic that is below the mean observed for the measure amongst the parents. Conversely, parents with a lower value of the characteristic have children with values higher than the mean amongst the parents. This is why the law was called "regression towards the mean". Curiously, the statistical method of line adjustment by the method of least squares is still called 'linear regression' today by one of Galton's followers, Pearson. The r value, which shows how well the experimental points have formed into a line, is **Pearson's linear regression coefficient**.

The results and the apparently antagonistic interpretations led to a scientific dispute that spanned the first decades of the 20th century. That dispute was significant in the discussion regarding the process of biological evolution, since **Charles Darwin**, one of the creators of the theory of evolution by natural selection, together with Alfred Russell **Wallace**, also English, believed that natural selection was a process that occurred on genetic variation in a continuous fashion, and was therefore a gradual one.

4.3 From Karl Pearson to Ronald A. Fisher

Modern statistics appears in the middle of the 19th century. One can say that this new stage in the development of statistics originated in biometric research laboratories.

We shall first look at Karl **Pearson**, a British mathematician.

He founded "*Biometrika*" (a very well known international magazine on biometry) and follower of Francis **Galton**. Known as the creator of applied statistics, he studied in Cambridge University and initially devoted his studies to hereditariness, applying statistical methods and developing Galton's theory. The work of Karl **Pearson** comprises an enormous quantity of publications, mainly in *Biometrika* magazine, which he founded together with Walter **Weldon** and Francis **Galton**.

Biometrics: A field of biology that applies statistical methods and probability calculation to the study of living beings.

He developed the theory of regression and correlation applied to the problems of hereditariness, he created the "**chi-square**" test and was one of the champions of the idea that statistics should be recognised as an independent subject and taught at the secondary education level (*Galeria dos Matemáticos*, 1991). He created the 'moment method' and the system of 'frequency curves', which are still used today to mathematically describe natural phenomena. The **Pearson distribution**, which is better known as the "**chi-square**" distribution (**χ^2**), constitutes the basis of small sample statistics of normal populations, is used to measure the confidence level of statistical results, tests hypotheses, etc.

Karl Pearson was born on 27 March 1857 in London. He is considered to be the creator of applied statistics. He was a follower of Francis Galton in his work on hereditariness. Despite the fact that all his work was in the field of biology, his great contribution to statistics resides in the discoveries he made in order to explain biological problems related to evolution and hereditariness.

Another important mathematician to the development of modern statistics was the Englishman William Sealey **Gosset**, better known as **"Student"**. **Student** worked as a chemist at the Guinness Brewery. It was there that he started to perform a range of experiments related to the control of the beer's quality. In the beginning **Student** used the normal distribution, but soon he experienced difficulties with the use of the "**Law of Error**" with small samples. To solve this problem he corresponded with the great statistician of the time, Karl **Pearson**, who had already developed ideas that led to the χ^2 distribution, but, like all other statisticians of the time, he was more interested in large samples. Nevertheless, **Student** developed the ***t*-test** and the results were published in *Biometrika* magazine.

He used the pseudonym **Student**, since the Guinness Brewery did not want competitors to know the statistical methods they used to improve the quality of their beer. Despite the great importance of his discovery, his work was largely ignored, only to be rediscovered later by **Fisher**. The *t* distribution is a distribution of theoretical probability that is similar to the reduced normal curve, differing from this through the use of a parameter called the **degree of freedom**. Degrees of freedom can be any real number greater than zero.

William Sealey Gosset was born on 13 June 1876 in Canterbury, England. He studied chemistry and mathematics and his contribution to statistics was the *t student* distribution. The factory where he worked did not allow him to publish the work under his own name, for they were afraid that competitors might be able to discover the level of quality of their products, so Gosset became known as "Student", which was the modest pseudonym used by this great statistician.

We shall now look at the work of the great statistician **Ronald A. Fisher**, one of the founders of modern statistics.

He became interested in the theory of evolution and selection, especially in the field of genetics, just like Francis **Galton.** His interest in the field of statistics arose from his work in genetics, which was the field in which most of his work was concentrated. He corresponded regularly with his great friend **Student**, and this led to **Fisher** making the distinction between **sample mean** and **population mean**. His area of interest was relatively small samples and not infinitely large ones. He was someone that did not like making mistakes and suffered immensely when he had to admit to an error. For this reason, he came up with various theories that he later tried to develop with others. He was rejected for national service in the First World War due to his poor eyesight and so he started to teach in a secondary school as a form of community service.

"… despite the fact that there is always uncertainty in statistics, it does not mean that there is a lack of precision - uncertainty can be the target of quantitative precision. Fisher did much to give form to this idea and make it a reality".
-- G.A.Barnard (Professor at Essex University)

It was at this time that his work in the field of statistics came to the attention of Karl **Pearson**, the most famous statistician of the time. **Pearson** criticised **Fisher's** work, maybe due to jealousy, which greatly hurt Fisher. This led to a great conflict between the two statisticians as both started to point out the errors that the other had made.[2] In 1919 he received two job offers: go to work in England with **Pearson** or to Rothamsted Agricultural Research Station. As he was not very friendly with **Pearson** he opted for the latter, about which he was also full of enthusiasm since the research station had observation records that went back almost one hundred years. He started analysing this data and he introduced a new set of methods such as

[2] http://www.mrs.umn.edu/~sungurea/introstat/history/w98/RAFisher.html

maximum probability (he studied all of its properties), variance analysis, hypothesis testing and experiment planning.

Ronald Aylmer Fisher was born on 17 February 1890 in East Finchley, London. He is considered to be one of the fathers and founders of modern statistics. He graduated in astronomy from Cambridge University, having been interested in mathematics from a very early age. His contribution to the evolution of statistics is primarily based on experiments performed at Rothamsted Agricultural Research Station. He developed a number of statistical methods in his work there, such as maximum probability, variance analysis, hypothesis testing and experimental planning.

These ideas have given researchers many instruments to handle variables, small samples and very precise estimates.

Fisher received three medals from the Royal Statistical Society: the Royal Medal (1938), the Darwin Medal (1948) and the Copley Medal (1955), and he was awarded a knighthood by Queen Elizabeth II in 1952.

He never stopped studying genetics and even theoretically proposed the existence of two new antibodies in his assessments of blood types. All of these statistical methods are studied in nearly all university courses and they have become part of everyday life.

4.4 Andrei Nicolaevitch Kolmogorov

Born on 25 April 1903 in Tambov, Russia, **Kolmogorov** showed an interest in mathematics from a very early age. When he was five or six, he discovered that the successive addition of odd numbers is equal to the successive square of whole numbers.

$1=1^2$
$1+3=2^2$
$1+3+5=3^2$
$1+3+5+7=4^2$
…
$1+3+\ldots+(2n-1)=n^2$

At school, **Kolgomorov** was the only child that invented mathematical problems, many of which were published in the school newspaper.

As referred to in Chapter 3, **Kolmogorov** set down the axiomatic basics of probabilities and developed a theory that constituted an enormous step forward in the field, creating a historic milestone. **Kolmogorov's axioms** basically establish that:

The Axioms of Probability

- A sample set and an algebra of events is always associated to the possible outcome of a random experiment;
- There is a non-negative number (greater than or equal to zero), called ***probability***, for all algebra events and this can be attributed to said event;
- The probability of the sample set is equal to 1;
- With regard to two disjointed events (that do not share any outcome) the probability of union is equal to the sum of their probabilities;
- The previous Axiom is true for infinite unions, as long as all event pairs are disjointed.

The application of mathematical logic to the principles stated above leads to the following fundamental properties of probability:

Fundamental Properties of Probability:

- The probability of any event is always greater than or equal to zero and less than or equal to one;
- The probability of an impossible event is zero;
- If the occurrence of an event requires the occurrence of another, then the probability of the first is less than the probability of the second event;
- The probability of the union of two events is equal to the probability of the first plus the probability of the second minus the probability that the two events will simultaneously occur.

4.5 The Twentieth Century

4.5.1 The Cradle of Statistical Applications

Statistics can be applied to nearly all areas of human activity. In **agriculture** Fisher made a great contribution due to his work at Rothamsted Agricultural Research Station. Statistical analysis methods permit the improvement of productivity, increased efficiency, the meticulous and methodical study of production conditions, as well as a host of other applications. "**Industrial applications** first appeared in the 1930s: control cards and lot control (which were the basis for the development of hypothesis testing) are maybe the first contributions that statistics made to the technological improvement of the industrial society. In the field of medical applications, the study of the effectiveness of pharmaceutical products, treatment quality and the

detection of possible causes of diseases are some of the applications of statistics" (Oliveira, 1995). The **State** needs to know the dimension of its population and so it uses statistics, in particular censuses, to make governmental decisions, e.g. if it knows how many people between 15 and 18 years of age live in a certain location then it will know if it needs to construct a secondary school there or not. **Meteorological services**, which are so important to air and sea transport, essentially provide statistical information. **Information technology** also uses statistics, in artificial intelligence and the assessment of computer network performance, for example. **Medicine** uses statistics to forecast certain diseases and to ascertain the effects that a certain medication may have on specific patients. In **Engineering** statistics is used mostly in quality control, for obtaining the percentage of defective output of a machine, for example.

4.5.2 Exploratory Analysis of Data

The classical techniques of statistics were designed to be as good as possible, based on a set of rigid assumptions. Experience and later investigation have shown us that the classical methods behave deficiently when the real situation moves away from the ideal situation defined in the set of assumptions. Recent developments, such as robust methods and exploratory data analysis, have contributed to improving the effectiveness of statistical analysis.

The main objective of exploratory analysis is to extract information from data, establishing relations between objects and variables. Exploratory analysis does not establish models *a priori*, but permits that relations observed in the data serve as the basis for hypotheses and proposed models.

There are two phases in data analysis: the exploratory phase and confirmatory phase. Exploratory data analysis highlights the flexible search for clues and evidence, while confirmatory analysis focuses on the assessment of the available evidence.[3]

4.6 - Future Trends

Statistical information is currently obtained, classified and stored on magnetic media and then made available through different information systems that provide intelligible information to researchers/the general public and society's organisations that is necessary to the development of their activities. The growth in the acquisition, storage and dissemination of statistical information, which has been extensively aided by the use of computers, has been accompanied by the rapid development of new statistical techniques and methods for data analysis.

Data engineering is a new area that has been significantly advanced by information technology.

[3] Análise Exploratória de Dados. Técnicas Robustas (1983)

The discovery of computer-based calculation has led to the development of algorithm families to handle data. These can be grouped together into the area of data mining.

5. Statistics in Portugal

5.1 Portugal and Statistics: Figures and History

"Statistical application in Portugal started, just like in other European countries, with the need for the state to better know the characteristics of its population. From the 16th century onwards, factors such as the implementation of the absolutist state, the development of the central administration and an increasingly broader and dynamic market caused an increase in demand for quantitative information to an extent where it became decisive.
There was the need to assess wealth, in terms of men and goods, during times of mercantilism, estimating the scope of military potential, assessing tax resources and sketching out state budgets (Sousa, 1995).

Country

Portugal, located in the southwest portion of the Iberian peninsula, is one of Europe's smallest countries. The country is rectangular in shape and is bordered to the north and east by Spain. Portugal's land border often follows the course of rivers, but mostly there are no natural barriers forming its borders. The border is the oldest in Europe, dating back to 1297.

According to the work "História da Estatística em Portugal" [The History of Statistics in Portugal] (Fernando Sousa, 1995), the register of events, particularly the counting of military forces, assets, income and expenditure, are the areas of greatest focus in the Middle Ages in Portugal. This era is noted for the scarcity of quantitative statistical data.

The need to quantify society arose from the King's need to know information regarding his army and the population to defend. The first known records concern *besteiros* (crossbow-wielding soldiers), who were subject to control lists. Later on, the quantitative relationship between the number of *besteiros* in each county (*conto*) and the respective population was established. The Church in the Middle Ages also produced numerous documents (censuses and property registers) regarding the socio-economic situation of the areas under its control. The crises of the 14th and 15th centuries required that church and secular landlords make better use of their properties,

which caused them to draft systematic inventories of assets and income and create land registers (*tombos*). These registers not only provided information to each landlord and let them better control the economic situation, but they also permitted income for each year to be forecast.

Condado Portucalense instituted in 1095 by King Afonso VI of Leon

Surveys were carried out by Portuguese monarchs, which investigated the status of royal rights and the legitimacy of the nobles' possession. These inquiries also allowed conclusions to be drawn on professional and economic organisation, as well as detect certain degrees of social stratification. King Dinis, based on the results of one such survey, ordered a general register, i.e. , a written record, to be drawn up so as to prevent ambitious nobles from seizing land and rights that were not rightly theirs. There were protests and complaints of course, and even an attempted revolt, but the King's will and orders prevailed.

As the Liberal State approached and the concept of nation based on a central administration became stronger, the call for the generalised statistical cover of the entire country began to be made, since a government cannot effectively administer uncertainty and the unknown. Plans emerged for kingdom-wide registers and systematic counts of houses were performed and the first all-encompassing statistical series on foreign trade commenced – General Trade Balance of the Kingdom of Portugal, 1776-1831. We can take these developments as being the symbol for the start of a new era. Statistical tables in a wide range of social areas appeared and overall figures were highlighted. The information, however, is still largely scattered, collected second-hand, produced by third parties, and not always complying with statistical quality and accuracy requirements – for example, parish priests were ordered to provide data on the population. Data was published on foreign trade (1842), municipal taxes (1845), the flow of goods through customs in Lisbon and Oporto (1856-1857). In the demographic field, the first census worthy of the name was performed in 1864. In other areas, *Annuario Estatistico* [Statistical Annual Report] was published in 1875, followed by independent series for other sectors (taxes, bank movements, transport, etc.).

House counts: count of the number of homes (households) in order to collect data to levy taxes and for military recruitment.

The entry into the era of statistics occurs gradually during the 19th century, with the creation of bodies that represent Portugal in different international conferences.

The effective employ of the collected data only really commences in the 20th century when statistics is developed as an applied branch of mathematics that is linked to the calculation of probabilities, which will assure the regular provision of summarised indicators, the sequential outlook of development trends, the possibility of forecasting. This, however, only becomes possible with the creation of Instituto Nacional de Estatística (INE) [National Statistics Institute of Portugal], in 1935 (Fernando Sousa, 1995).

Important and known statistical works after the creation of Portugal and prior to the founding of INE:

- Census of Crossbow Men per County, King Afonso III (1260-1279);
- Census of Crossbow Men per County, King João I (1421-1422);
- Home Count or General Register of the Kingdom, King João III (1527);
- Inventory of Soldiers, King Filipe III (1639);
- List of Homes in Portuguese Lands, King João V (1732), also known as the Census of the Marquis of Abrantes;
- Pina Manique Home Count, Queen Maria I (1798);
- General Census of the Kingdom, King João VI, also known as the Census of the Count of Linhares (1801);
- General Censuses of 1835 and 1851.

5.2 Censuses in Portugal

The word ‘census’ derives from *Censere* which means ‘to tax’ in Latin.

The first Portuguese census was performed between 31 December 1863 and 1 January 1864. This census was founded on the guidelines of the first international statistics conference, held in Brussels in 1853. Censuses had already been held in Portugal before this, as noted above, but since they were not exhaustive and/or were not based on credible statistical principles, they cannot be deemed to be equivalent to those commenced in 1864.

The method of direct data collection was chosen for these censuses, with the population being surveyed on the same day and at the place they had spent the night. Censuses should have been

The first census held in Portugal according to the standards of the international statistics conference was in 1864.

performed every 10 years from then onwards, but the next census was in 1878, which was followed by one in 1890. From 1890, censuses have been regularly performed, with a few exceptions, at 10-year intervals.

In 1940, and thereafter, the censuses have been performed by Instituto Nacional de Estatística. From 1970 onwards a general census of housing has been performed at the same time.

Fourteen censuses of the population and four housing censuses have been performed to the present day.

The following table contains all of the censuses performed in Portugal, and a summary of the preceding statistical history:

1864 – 1 January (1st General Census of the Population)	
	The first general census of the population is performed, based on the guidelines of the first international statistics conference that was held in Brussels in 1853.
1878 – 1 January (2nd General Census of the Population)	
	The second general census of the population is performed. It was more comprehensive than the previous one, in terms of variables studied and groupings, though, compared to modern standards, its content is quite reduced.
1890 - 1 December (3rd General Census of the Population)	
	This census was performed based on new methodological guidelines, in accordance with the international statistics conference held in St. Petersburg in 1872; the collected data describing the population and families was much more comprehensive.
1900 - 1 December (4th General Census of the Population)	
	The data collection methods, data handling and presentation was quite similar to the previous census, with a few innovations.
1911 - 1 December (5th General Census of the Population)	
	The methodology and variables analysed remained unchanged.
1920 - 1 December (6th General Census of the Population)	
	The methodology and variables analysed remained unchanged.
1930 - 1 December (7th General Census of the Population)	
	There were no great alterations to the characteristics analysed, though economic variables continued to be poorly encompassed.
1940 - 12 December (8th General Census of the Population)	
	This was the first census performed by Instituto Nacional de Estatística and it is deemed to be a historic milestone in Portuguese censuses. A new implementation method was used. More accurate economic data is collected and economic information is deemed to be an important part of the census.
1950 - 15 December (9th General Census of the Population)	
	The methodology of the previous census was followed, with a few improvements, such as the improved technique with regard to closed-ended questions.
1960 - 15 December (10th General Census of the Population)	
	Past data is published for the first time. The content of the 1950 and 1960 censuses is very similar to that of the 1940 census.
1970 - 15 December (11th General Census of the Population) (1st General Housing Census)	
	The first general housing census is performed. However, the audacious plan, which aimed to meet the government's numerous requests for information, was not an executive success, particularly in relation to the overall results to be published.
1981 - 16 March (12th General Census of the Population) (2nd General Housing Census)	

	The censuses of the population and housing closely followed international recommendations (EEC/UN). The concepts are meticulously applied in nearly all areas and the data collected is significantly unbundled in geographical terms.
1991 - 15 April (13th General Census of the Population) (3rd General Housing Census)	
	The methodology used was the same as in the previous census, with the development of some of the data preparation and handling techniques that had been instigated in 1981. A Spatial Referencing Geographical Database was built, which comprised a set of cartographical charts containing information permitting the division of parishes into statistical sections and sub-sections.
2001 - 12 March (14th General Census of the Population) (4th General Housing Census)	
	The major difference concerns the technologies used (digital mapping, the use of geographical information systems, optically readable questionnaires, computer-assisted coding and the improvement of the automatic correction of incoherent answers). A new question concerning disability is added to the individual questionnaire.

5.3 The Teaching of Statistics in Portugal

Statistics in Portugal is a branch of applied mathematics, as it is in many other countries. Its study and development as a science has grown as society has progressed and nowadays statistics is present in nearly all fields of knowledge.

As João Branco (JME-190) states, at the end of the 19th century there was the generalised emergence and acknowledgement of problems of a statistical nature in the fields of science, in industry and in government, which led to the growth of interest in statistics. The speed with which these developments occurred generated a crisis regarding the shortage of skilled personnel with knowledge of statistics, as such personnel were very much sought after by the many different institutions that wished to make use of the new methodology to further their activities. This fact generated the need to teach statistics to an increasingly greater number of people. Priority was initially placed on teaching at an advanced level in order to improve the techniques of those interested in the profession or those involved in scientific research. The teaching of elementary statistics was only focused on at a later date. This level aimed to supply the basics to natural science students and students interested in entering the statistics profession. This basic knowledge, despite its elementary nature, also began to be taught in postgraduate courses or during the final years of undergraduate courses. However, it was soon concluded that these courses in elementary statistics had to be provided earlier on in a student's academic development, i.e. more towards the beginning of undergraduate courses.

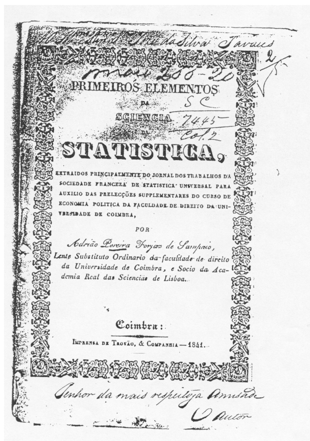
5.3.1. Statistics in Secondary Education

According to João Branco (JME-190), the teaching of statistics at secondary level arose when it was acknowledged that the general public needed to be supplied with a coherent system of

statistical ideas and the capability to be able to innately use these ideas in a society that was becoming evermore based on numerical information and data. A meeting of great import to the development of the teaching of this discipline was held in Royaumont in 1959, with the backing of the European Organisation for Economic Cooperation (OECE), an organisation that was replaced by the Organisation for Economic Cooperation and Development (OECD) in 1961. Mathematicians from all over the world attended this event, with the purpose of studying a profound reform of the teaching of mathematics at pre-university level. They concluded that the teaching of probability calculation and statistics should be added to secondary education syllabuses.

The general movement to modify mathematics syllabuses and teaching methods in secondary education that started to emerge also appeared in Portugal, particularly through the publications and meetings produced by the OECD. José Sebastião e Silva, one of Portugal's most important mathematicians, was charged with the responsibility of modernising the teaching of mathematics at the upper secondary level. The alteration of syllabuses, to adapt them to the scientific and technological revolution occurring at that time, led to the inclusion of a number of mathematical areas in Portuguese secondary education for the first time, such as probability calculation and statistics. In 1963-64 the first three classes were created and implemented on a trial basis. This trial, which was successively repeated over a number of years and was expanded to include dozens of different secondary schools throughout the country, laid the groundwork for the inclusion of these areas in the national syllabus of secondary education.

The movement that led to the inclusion of statistics in secondary education arose at exactly the same time as definitive steps were taken to include statistics in mathematics undergraduate courses. The movement for the teaching of statistics in higher education was particularly active in the Faculty of Science of Lisbon University. The first undergraduate course in Probability and Statistics was created in 1982. A principal figure of the movement was José Tiago de Oliveira, a great scientist who developed a passion for statistics and its problems at all levels, including its teaching at secondary level (JME-190).



Front cover of the work of Professor Adrião Sampaio

According to Adrião Cunha (2001), statistics was first taught in Portu of Law of Universidade de Coimbra [Coimbra University]. The person responsible for the inclusion of the subject was Professor Adrião Sampaio, through his work entitled *Primeiros Elementos da Ciência Estatística* [Basics of Statistical Science] which he used as a source book in the supplementary lessons he gave in the course of political science of which he was in charge in the abovementioned faculty.

5.4 INE and the National Statistics System

In Portugal, INE is the operational body responsible for the collection, handling, analysis and dissemination of the country's national statistics. There are, nonetheless, bodies that manage to allocation of duties of the entire statistical process.

The National Statistics System (SEN) is composed of Conselho Superior de Estatística (CSE) [Higher Statistics Board], Instituto Nacional de Estatística - INE [National Statistics Institute of Portugal] and Delegated Bodies of INE. The principles of SEN are exclusivity, technical independence, professional secrecy regarding statistics, statistical authority and the cooperation of public services.

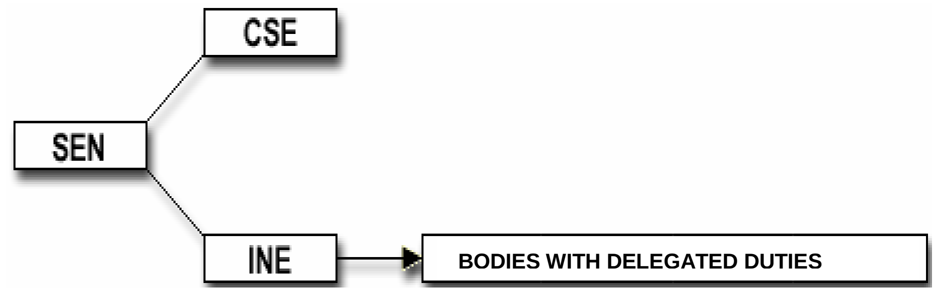
The Higher Statistics Board (CSE) is the state entity that supervises and coordinates SEN.

Composition of the National Statistics System

Under Law no. 6/89 of 15 April, SEN is composed of the Conselho Superior de Estatística (CSE) [Higher Statistics Board] and Instituto Nacional de Estatística (INE) [National Statistics Institute of Portugal]:

- CSE is the state body that supervises and coordinates the National Statistics System (SEN) (Article 8 of said law);
- INE is exclusively responsible for the notation, analysis, coordination and dissemination of official statistical data (Article 3 of said law).

However, Article 19, paragraph 1 of said legislation states, “(…) to implement its duties, INE can delegate official functions of notation, analysis and the coordination of statistical data to other public services”. The dissemination of statistical data is excluded from this delegation of duties - INE remains exclusively responsible for this task.



Instituto Nacional de Estatística (INE) [National Statistics Institute of Portugal] was created in 1935 to meet ever increasing requirements for statistical information. Its objective is to perform surveys, censuses and other statistical operations; create, manage and centralise statistical unit files; access data on individual entities (except for data on private individuals) that is made available by entities managing public services; perform pure and applied statistical studies and socio-economically analyse available statistical data; foster the training of SEN personnel and cooperate with foreign statistics organisations.

INE became a public institute in 1989, and was granted legal personality, administrative and financial autonomy, and its own assets.

INE currently produces dozens of official publications, not only in demographical fields but also in many other fields of application, such as industry, trade, education and many others. We all have the right to access these publications and contribute to the construction of the same.

See also…

Publications

- BÉDARIDA et al. (1987), *Pour une Histoire de la Statistique*, Economica.
- CUNHA, Adrião Simões Ferreira (2001), *Nótulas Históricas em Torno do Sistema Estatístico Nacional* [Historical Notes on the National Statistics System], Lisbon, Instituto Nacional de Estatística.
- DAVID, F.N. (1998), *Games, Gods and Gambling, A History of Probability and Statistical Ideas*, Dover Publications, Inc. Mineola, New York.
- *Galeria dos Matemáticos do Jornal de Matemática Elementar (2º Volume)* [The Gallery of Mathematicians of *Jornal de Matemática Elementar* [Journal of Elementary Mathematics - Volume 2], (1994), Lisbon.
- *Galeria de Matemáticos do Jornal de Matemática Elementar* [The Gallery of Mathematicians of *Jornal de Matemática Elementar* [Journal of Elementary Mathematics], (1991), Lisbon.
- HEYDE, C.C., SENETA, E. (2001), *Statisticians of the Centuries*, Springer, New York.
- HOAGLIN, David C., MOSTELLER, Frederick, TUKEY, John W. (1983), Novas Tecnologias/Estatística: *Análise Exploratória de Dados. Técnicas Robustas* [Exploratory Data Analysis - Robust Techniques], Edições Salamandra.
- INE, *Programa Global dos Censos 2001* [2001 Global Census Programme], Instituto Nacional de Estatística, Lisbon, available at: http://www.ine.pt/censos2001/Organizacao/programaglobal.asp
- JME - Jornal de Matemática Elementar nº190, [Journal of Elementary Mathematics, no. 190], Lisbon.
- LOUREIRO, Cristina; OLIVEIRA, Fernanda; BRUNHEIRA, Lina, *Ensino e Aprendizagem da Estatística* [Learning and Teaching Statistics], Sociedade Portuguesa de Estatística [Portuguese Statistics Society], Associação de Professores de Matemática [Association of Mathematics Teachers], Departamento de Educação e de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa [Department of Education and Statistics and Operational Research of the Faculty of Science of Lisbon University] 2000.
- OLIVEIRA, J. Tiago (1981), O Ensino Iniciário da Estatística. *Actas do II Colóquio de Estatística e Investigação Operacional* [The Initial Teaching of Statistics - Minutes of the Second Symposium on Statistics and Operational Research], Fundão.

- OLIVEIRA, J. Tiago (1995), *Collected works* (Volume II), Évora.
- PERSON, E. S., KENDALL, Sir Maurice (1820), *Studies in the History of Statistics and Probability*, volume I, Charles Griffin & Co Ltd, London.
- SOUSA, Fernando (1995), *História da Estatística em Portugal* [The History of Statistics in Portugal], Instituto Nacional de Estatística, Lisbon.
- STIGLER, Stephen M. (1986), *The History of Statistics, The Measurement of Uncertainty before 1900*, Belknap Harvard.

Websites:

http://www.sobiografias.hpg.com.br (a number of biographies of historical figures);

http://users.hotlink.com.br/marielli/ (this mathematics site contains a number of biographies of famous mathematicians as well as the history of numbers, arithmetic, etc.);

http://www.educ.fc.ul.pt/semtem/semtem99/sem21/framegeral.htm (the history of Pascal's triangle);

http://www.mala.bc.ca/~johnstoi/darwin/sect4.htm
http://www.amostraestatistica.hpg.ig.com.br/historia.htm
http://www.esgb-antero-quental.rcts.pt/NMAT/m
(these sites contain parts of the history of statistics);

http://www.mat.uc.pt/~bebiano/Atractor/esta.htm (some mathematical models, including the Quincunx);

http://www.geocities.com/g10ap/matematicos/os_grandes_genios.htm (biographies of great mathematical geniuses);

http://www.ib.usp.br/evolucao/QTL/historiaqtl.html (includes an explanation of Galton's law of regression towards the mean)