ADOLPHE QUETELET, FATHER OF MODERN STATISTICS?

Thomas S. Graves

BU-581-M May 1976

Abstract

A brief history of statistics up until 1820 is presented, along with a quick sketch of the development of the normal distribution and the laws of large numbers before 1820. Then a biography of Adolphe Quetelet is given and it is shown how he combined the art of statistics with the theory of probability to form the science of statistics. His other contributions to statistics are also listed.

Introduction

The purpose of this paper is twofold. First a brief history of statistics up to 1820 will be given, along with a quick sketch of the development of the laws of large numbers and the normal distribution. Secondly, a biography of Adolphe Quetelet with the emphasis on his achievements in statistics will be presented. It will be shown how he combined what was then called statistics with probability theory to develop what we today call statistics. For this reason, Quetelet is often referred to as the father of modern statistics. He was also influential in the standardization of international statistics, helped to promote the use of statistical methodology in other fields of science which rely upon observational data and laid the groundwork for modern sociology. He changed the role of the statistician from that of artisan, gathering data and making official tables to that of scientist, seeking out the laws of nature.

History of Statistics

The practice of recording the numerical quantities which make up the state goes back to the very origins of recorded history. The Greeks, Persians, Egyptians and Romans all took censuses to determine the strengths and weaknesses of their respective political states. The Bible, in Samuel II, gives an estimate of the world population of 3,800,000.[1] Meitzen, in his book History, Theory and Techniques of Statistics (1891) gives many such examples.

In the 15th century, with the rise of centralized monarchies, a new need for knowledge of these states also evolved. A monarch, in order to stay powerful, needed to know such things about his domain as the number of taxable citizens, the number of draft-age males, the distribution of his population and imports and exports, along with a similar knowledge of the domain of his neighbors and potential enemies. Birth and death counts were also kept for the first time on a wide basis. The most efficient of these state evaluators, as one might guess, were the Germans who created the first school of statistical thought. A typical work of this school was the Cosmographia of Sebastian Muenster (1489-1552) which contained maps, state organization and military capacity of the countries of Europe. Statistical studies first made their appearance in a university in 1660 at Helmstedt in the lectures of Hermann Conring (1606-1681). Much of what he called statistics would today be called social studies. The word "statistik" was first used by Gottfried Ackenwall (1719-1772) who derived it by combining two Italian words reagione di stata, meaning practical politics and statisto, meaning statesman and defined it to mean the study of the life of the state with a view to ascertaining its strengths and weaknesses as a guide for the statesman.[2] Achenwall put these facts about the state into a concise, easily read form and was considered at that time to be the father of statistics. The followers of Achenwall did not bother themselves with enumeration of data, but were content with giving verbal description and actually looked down upon what they called "table statisticians". However, distrustful bureaucrats forced them to use tables. Opposed as the German statisticians were to the use of tables in their statistical writings, they were still better off in modern terms than the French, whose idea of conducting a census until 1817 was to ask the "experts" what the figures should be.

About the same time that Conring was delivering his lectures, another group was taking shape, made up primarily of Englishmen concerning themselves with the study of populations as such and using different methods and concepts than the German "statists". The first of this group was Captain John Graunt (1620-1674), a military man turned shopkeeper who threw some light on the regularity of social phenomena when he presented to the Royal Society in 1662 his "Natural and Political Observations upon the Bills of Mortality with Reference to the Government, Religion, Trade, Growth, Air, Disease and the Several Changes of the City of London". The Bills of Mortality of reference were weekly newsletters which listed the number of births and deaths in London, with cause of death. They were of particular interest during plague years. Graunt came up with some surprising results, e.g. that 14 boys are born for every 13 girls and he gave a mortality table which showed that of every 100 persons born, 36 die before age 6, 24 between 6 and 16, 15 between 16 and 26, and successively 9, 6, 5, 4, 2 and 1 in the next decades.

He also made some crude attempts at estimating the population of London. This was really the first time that anyone tried to make inference from data. William Petty (1623-1687), another Englishmen, carried on with his work "Political Arithmetick ...", where he attempted to use Graunt's methods to obtain information about the rest of Europe. In it, he complains about the lack of available data. Desiring to reach the truth through calculation, and lacking observations to calculate with, he arrived at some quite amusing results. His major contribution to statistical thought was his belief that one must use quantitative entities to obtain knowledge. He defines political arithmetic as "the art of reasoning by figures upon things in government."[3]

It was also around this time that insurance was becoming big business. Johan De Witt (1625-1672), a Dutch legislator, was in 1671 the first to attempt to find a scientific basis for annuity tables; i.e., deriving an annuity table from a mortality table. However, having no data to work with, he made up his mortality table from assumptions. He assumed a constant chance of death from ages 4 to 53, 3/2 chance from 53 to 63, double chance from 63 to 73 and triple chance thereafter. Because of this, De Witt became the Grand Pensionare of Holland, but due to his various political intrigues, he and his brother were lynched by an angry mob.

Edmund Halley (1656-1742), the astronomer of Halley's Comet fame, published in 1693 the first complete tables of mortality based on data obtained in the cities of London and Dublin.

The man who brought this whole so-called school of Political Arithmetic together was Johan Süssmilch (1707-1767), a German military chaplain who based his works on those of Graunt, Petty and others. In his book, he made some good (by today's standards) estimation of population parameters, such as the numbers of births, deaths and marriages per year for his native city of Breslau, and some distinct advances in statistical methodology. He realized that 1) social phenomena have causes, 2) the regularities existing in his statistical results revealed the rules of the existing social order, and 3) constancy of results can only be obtained with large sample sizes. His results pleased him, as they showed "the rules of order which God's wisdon and goodness have established".[4] Further studies in this area were conducted by Malthus and Fourier, both of whom Quetelet knew personally. Malthus' "Essay on Population" gave worldwide stimulus to the study of population, and Fourier developed some excellent tables of mortality and population.

We see then that the school of political arithmetic of Graunt through Süssmilch, in contradiction to the descriptive school of Achenwall, began by laying emphasis on the method of inquiry rather than description of results. It was up to Quetelet to combine the two with his studies of causation and correlation.

Histories of the Normal Distribution and the Laws of Large Numbers

The histories of the normal distribution, the laws of large numbers and the central limit theorem are all inextricably intertwined. The normal distribution is at the heart of the modern central limit theorem, which wasn't discovered until after the first attempts had been made to form a law of large numbers. The normal distribution is that probability distribution defined by the density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-u}{\sigma})^2}$ where u and σ^2 are, respectively, the mean and variance of the normally distributed random variable.

This result doesn't seem very important until one sees the central limit theorem, which says in effect: If the outcome of a process is the additive result of a large number of independently acting random factors, no one of which dominates the course of the process as a whole, then regardless of the nature of the probability laws governing these independently acting random factors, the probability law governing the process as a whole is approximately normal.[5] The laws of large numbers are a group of theorems dealing with events which depend upon a set of random factors whose size increases without limit and which will occur with probability as close to one as is desired. The central limit theorem is included in this group.

The first tract on probability theory was written by Christian Huygens (1629-1695) in 1657. It contained nothing of interest relative to our present topic, but it did pose several unsolved problems which stimulated more works, among them the _Ars Conjectandi_ of Jacques Bernoulli (1659-1705) published posthumously in 1709. _Ars Conjectandi_ is in four parts. They are: a reproduction of Huygens' text, with Bernoulli's solutions to the unsolved problems; a treatise on combinations and permutations; an application of combinations and permutations to gambling problems; and an application of the above topics to civil, economic and moral affairs. This last section was left unfinished, but it does contain the first attempt at a law of large numbers. Bernoulli writes:

> THE PROPOSED PRINCIPLE. To avoid the ennui of circumlocutions, I shall call those cases in which a certain event can happen _fecund_ or _fertile_, and _sterile_ those cases in which the same event can fail to occur. Moreover, experiments are _fecund_ or _fertile_ in which any one of the fertile cases is detected, and _infecund_ or _sterile_ in which any one of the sterile cases is observed to happen. Let, therefore, the number of fertile cases to the number of sterile cases be exactly or approximately in the ratio r/s, and so to the number

> of all the cases in the ratio $r/(r + s)$ or r/t, which ratio has the adjoining limits $(r + 1)/t$ and $(r - 1)/t$. It must be shown that so many experiments can be performed that it will be more likely by as many times as one wishes (say c times) that the number of fertile observations will fall within these limits than outside them; that is, that the number of fertile observations will have to the number of all observations a ratio which is not greater than $(r + 1)/t$ nor less than $(r - 1)/t$. [6]

In other words, by increasing the sample size, one can make the probability of r/t falling within the interval $[\frac{r-1}{t}, \frac{r+1}{t}]$ as large as one wishes.

The next major advance was that of Abraham DeMoivre (1667-1754), a French protestant expatriated to England in 1688 due to the revokation of the Edict of Nantes in 1685. His major work, Doctrine of Chances published in 1738, is essentially a gambler's manual which gives solutions to problems of amount of wager and advantage of player. His version of the central limit theorem was first published in a pamphlet in 1731 and later included in Doctrine of Chances. There he proposed an approximation of the binomial distribution with a curve of the form e^{-t^2}. He showed that for y_o being the largest (middle) term of the expression $(1+1)^n = \binom{n}{o} + \binom{n}{1} + \cdots + \binom{n}{n}$, y_t a term which is t terms away from y_o, and n very large, that the following two equations held:

$$\frac{y_o}{2^n} = \frac{2}{\sqrt{2\pi n}}$$

and

$$\ln \frac{y_t}{y_o} = \frac{-2t^2}{n}$$ [7]

which together imply

$$\frac{y_t}{2^n} = \frac{2}{\sqrt{2\pi n}} e^{\frac{-2t^2}{n}} .$$

From this, DeMoivre could calculate $\frac{y_o + y_1 + \cdots + y_t}{2^n}$ by solving the integral $\int_o^t \frac{2}{\sqrt{2\pi n}} e^{\frac{-2X^2}{n}} dX$, which he could approximate with a series expansion.

DeMoivre, like Sussmilch and a surprising number of other probabilists and statisticians, was also wont to connect his results with theology:

> But such Laws, as well as the original Design and Purpose of their Establishment, must all be *from without*: the *Inertia* of matter, and the nature of all created Beings, rendering it impossible that any thing should modify its own essence, or give to itself, or to any thing else, an original determination or propensity. And hence, if we blind not ourselves with metaphysical dust, we shall be led, by a short and obvious way, to the acknowledgement of the great MAKER and GOVERNOUR of all; *Himself all-wise, all-powerful* and *good*. [4, pp. 251-252] 8

Karl Pearson, writing in 1926 had this comment:

> As DeMoivre appropriately observes, he has reserved to himself 'the right of enlarging my own thoughts.' That enlargement, developing Newton's idea of an omnipresent activating diety, who maintains mean statistical values, formed the foundation of statistical development through Derham, Niewentyt, Price to Quetelet and Florence Nightingale. ... The causes which lead DeMoivre to his "Approximatio" or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that the post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the eighteenth century -- in particular that of the scientists who were member of the Royal Society -- must remain obscure. [3, p. 552] 9

The next person to write on this general problem was Thomas Simpson (1710-1761), an Englishman who set himself to the task of proving the advantage of the practice of taking the mean of several observations. Simpson writes:

> ... that the method practiced by astronomers, in order to diminish the errors arising from the imperfections of instruments, and of the organs of sense, by taking the Mean of several observations, has not been so generally received, but that some persons, of considerable note, have been of opinion, and even publickly maintained, that one single observation taken with due care, was as much to be relied on as the Mean of a great number. [1, p. 82] 10

Simpson hypothesized two different probability distributions for the errors about the mean and showed that in each case the mean could be made to lie within certain limits with a specified probability. This marked the first time anyone had written that repetition of measurements gives rise to errors which follow a probability distribution. Simpson's work was elaborated upon by men such as LaGrange, LaPlace and Daniel Bernoulli, each of whom attempted the problem assuming different distributions for the errors. The first to hypothesize that the errors were distributed normally was Carl Friedrick Gauss (1777-1855) in his Theoria Motus of 1809, an assumption that he used for his principle of least squares. Since this principle worked so well in practice, the idea of errors of measurement in astronomy following a normal distribution (not yet named "normal") caught on. Adams puts it very well:

> One outgrowth of these developments was the emergence of the hypothesis of elementary errors -- the theory than an error of observation is the sum of a large number of elementary component errors which arise from different, independent sources, were each elementary error is negligible in comparison to the sum -- as an explanation for the normal law of errors. On the basis of this hypothesis it was argued that the probability law governing the error distribution arising from a measurement process is approximately normal, that is, the probability that an error arising from the measurement process is between two values, b and c, is approximately
>
> $$\frac{h}{\sqrt{\pi}} \int_b^c e^{-h^2x^2}\,dx$$
>
> where h is an experimentally determined constant whose value depends on the conditions under which the observations are made. This, from the modern point of view, is an intuitive face of the Central Limit Theorem. 11

LaPlace (1749-1827) generalized the central limit theorem to the case of errors of measurement in 1812 in his Theorie Analytique des Probabilités. He said that when the errors are governed by any arbitrary probability law having finite mean and variance, the mean of the errors is distributed normally for large sample sizes. It was this same LaPlace who taught Quetelet the theory of probability during Quetelet's stay in Paris in 1823.

Adolphe Quetelet (1796-1874)

Adolphe Quetelet was born on the 22nd of February, 1796 in Ghent, Belgium. His father, a well known municipal officer, died when Adolphe was seven. He attended the Lyceum at Ghent and graduated at 17 to become an instructor of mathematics at the New University at Ghent. There he earned his Ph.D. in mathematics under Jean Guilluame Garnier, professor of astronomy and higher mathematics. Until age 30, Quetelet also wrote poetry and composed an opera, but apparently Garnier's influence over him was quite decisive. Quetelet writes:

> Little by little his conversation, always instructive and animated, gave a special direction to my tastes, which would have led me by preference towards letters. I resolved to complete my scientific studies and followed the courses in advanced mathematics given by M. Garnier. 12

In 1819 he became chairman of elementary mathematics in the Atheneaum at Brussels. In 1820, he was elected to the Academic Royal des Sciences et Belles-Lettres de Bruxelles, a lethargic organization into which he was to infuse new life. He aroused a new interest in astronomy in Belgium and in 1823 was sent to Paris by the state to learn about the latest in astronomical hardware. There he met LaPlace, Poisson, Malthus, Fresnel and Fourier, men who would have a great influence on his later work. He studied probability under LaPlace and mathematics under Fourier. He returned to Belgium and in 1827 was charged by the king to buy the instruments needed to open an observatory. However, due to political haggling, the money was not appropriated until 1832. Until then, Quetelet traveled about Europe, visiting learned societies, observatories, and famous men. In 1829 he spent a week with Goethe and afterwards carried on an amicable correspondence. After the observatory was opened in 1832, Quetelet devoted his life to the academy, and astronomical and scientific research.

During this time Quetelet developed an interest in statistics. In 1825, he presented to the Academy his first statistical work entitled "Mémoir sur les lois des naissances et de la mortalité à Bruxelles". In 1826 he began arguing for a national census and in 1828 it was taken under his auspices. In 1833 he was a delegate to the meeting of the British Association at Cambridge. There he worked for the formation of a statistical committee which the next year became the Royal Statistical Society of London, of which he was a charter member. It was here that he met De Morgan, the British mathematician. De Morgan was at that time working on probability theory as applied to legal testimony, miracles, moral questions and jury verdicts and while not making any progress, his ideas had a strong influence on Quetelet.

In 1835, Quetelet published Sur l'homme et le dévelopment de ses facultés, ou essai de physique sociale, the forerunner of the modern sociology text where he presented his controversial concept of the "average man". In 1844 he was named as head of the Commission Centrale de Statistique, which took a second census in 1846 and every ten years thereafter. He helped to form the International Statistical Congress in 1851, chaired its first meeting in 1853 and there urged for uniformity of international statistics. He was the first foreign member of the American Statistical Association. He died on February 17, 1874 in Brussels.

Quetelet's contribution to modern statistics can be divided with some overlap into three classes: application of probability theory to statistics; introduction of new analytical techniques; and improved methods of collecting data.

Application of Probability to Statistics

It was already known, especially in astronomy, that if one took several observations of the same phenomena, such as the position of a particular star, these observations would not all be the same. If you looked at the deviations of these observations from their mean, you could see that these deviations or "errors" behaved as random variables following some probability distribution. Quetelet's contribution was the idea of looking at groups of observations of similar objects as if they were different measurements of the same object and seeing that their deviations about the mean followed a probability distribution. He first explains this phenomenon in Sur l'homme ··· in 1835, and later in his Letters Addressed to H.R.H. the Grand Duke of Saxe Colburg and Gotha on the Theory of Probabilities as Applied to the Moral and Political Sciences.

Quetelet demonstrates this idea with an example of taking chest measurements of Scottish soldiers:

> I now ask if it would be exaggerating,to make an even wager that a person little practised in measuring the human body would make a mistake of an inch in measuring a chest of more than 40 inches in circumference? Well, admitting this probable error, 5,732 measurements made on one individual would certainly not group themselves with more

> regularity, as to the order of magnitude than the 5,738 measurements made on the Scotch soldiers; and if the two series were given to us without their being particularly designated, we should be much embarrassed to state which series was taken from 5,738 different soldiers, and which was obtained from one individual with less skill and ruder means of appreciation. 13

He goes on to extend this idea of chest measurements deviating about an average to the idea of an "average man", about whom all men deviate in one way or another in all quantities, physiological, mental and moral. He explains that this concept is merely an abstraction, i.e. that there needn't be an "average man" alive and walking about somewhere on Earth, although he doesn't say that this is impossible. Many of his critics apparently overlooked this explanation, since much of their criticism centers precisely on this point. He also talks about each sex and race having their own average. He illustrates both points with an example:

> Let a desert island be peopled to-morrow, by placing upon it 1,000 men of the tallest race, the Patagonians for example, all six feet high, and 1,000 Laplanders only four feet and a half high, the mean height in this island will be five feet and a quarter and yet not one man will be of that height. Grouping them in order of size, we could form but two groups, and the law of possibility will be completely in fault,--it would in appearance at least be so. But we see at once that the difference only proceeds from the mixture of heterogeneous things,--men of different races, who have different laws of development. 14

He sees that this theory can have many useful applications. He finds an immediate example when he looks at the height of French conscripts, and by comparing the distribution with his normal approximation, he sees that there has been cheating going on at the draft physicals:

"The number rejected for deficiency in height is much exaggerated. Not only can we prove this, but we can determine the extent of the fraud."

If we compare the numbers observed with the numbers already calculated, we shall be able to give the following table:--

Height of Men	Number of Men		Differences of Results
	Measured	Calculated	
INCHES			
Under 61·821	28620	26345	+ 2275
61·821 to 62·884	11580	13182	- 1602
62·884 to 63·947	13990	14502	- 512
63·947 to 65·010	14410	13982	+ 428
65·010 to 66·073	11410	11803	- 393
66·073 to 67·136	8780	8725	+ 55
67·136 to 68·199	5530	5527	+ 3
68·199 to 69·262	3190	3187	+ 3
Above 69·262	2490	2645	- 155

If we regard the magnitudes of the numbers, it will be observed that the only remarkable difference between the numbers calculated and observed is in the smallest heights. The calculation shows that 2,275 men have been rejected for want of height, which appears ought not to have been the case according to the law of continuity. This number is found wanting in the two extreme categories, which are too small, the one by 1,602 men, and the other by 512 men, giving in 15
the total 2,114, a number little differing from the preceding.

So we see that of 100,000 conscripts, approximately 2,275 men cheated in the height exam to get out of the service.

The more controversial part of this theory was its application to mental and moral characteristics. In Quetelet's studies on crime and criminal behavior, he saw that the number of particular crimes committed from year to year was nearly constant. From this he postulated that all people had a moral quality which was distributed normally over the population about a mean, with people having crininal tendencies at one end of the distribution and people having an abhorrence of crime on the other. The apparent weakness in this argument was that critics interpreted him to mean that man has no free will, but is fated to be what he will be. Quetelet qualified this by saying that no man was forced into committing a crime by these laws, but that society could place him in a position where he was more likely to commit a crime than another man in a different position. In keeping with the laws of large numbers, given a large group of such people with criminal tendencies, a certain proportion would commit crimes. Quetelet was not the first to notice the regularity of voluntary actions. Apparently Condorcet and Madame de Stail also noticed this phenomena in the 18th century, but Quetelet was the first to give it a scientific basis. He also saw that birth, death, marriage and suicide rates tended to remain constant or have small variations which followed the same pattern as other key quantities, such as the price of grain. The constancy of these results led Quetelet to believe that it would be only a matter of time before the actual laws governing these occurrences were found, just as they had been in physics. For this reason, he named this study of human behavior "social physics".

Another application of probability to statistics was that, through distribution theory, Quetelet could prove that one obtained more precise results using large samples, a result at which previous statisticians could only guess. He defined several measures of precision, the principle one being his "scale of precision", which consisted of giving the limits in which one half of the observations lay, one fourth being on either side of the mean. He reported that the scale of precision varied as square root of n, where n was the sample size. This result was previously known to astronomers and probabilists, but no one had yet applied it to statistics.

Quetelet also noted that the distribution of deviation of biological phenomena about their mean, although usually symmetric, need not always be so, and he gives several examples where asymmetry may arise.

New Analytical Techniques

One of the consequences of the theory of errors of measurement as applied by the 18th and 19th century astronomers was the idea that there were causes for errors of measurement. Usually one could find many causes, such as aberration of the lens, atmospheric conditions and different astronomers making the measurements. Quetelet applied this idea to his statistics, found principally in chapter three of Letters ..., which is entitled "The study of causes".

Quetelet breaks causation into three categories: constant cause; variable or periodic cause; and accidental cause. He defines them as follows:

> Constant causes are those which act in a continuous manner, with the same intensity, and in the same direction.
> Variable causes act in a continuous manner, with energies and tendencies which change either according to determined laws or without any apparent law. Among variable causes, it is above all important to distinguish such as are of a periodic character, as for instance the seasons.
> Accidental causes only manifest themselves fortuitously, and act indifferently in any direction. 16

Compare this with the notation given in W. T. Federer's Statistics and Society, a modern beginning-level statistics text published in 1973:

> The total variation in measurements excluding mistakes or blunders may be written as
>
> total variation = assignable cause + bias + random error . 17

The causes which Quetelet refers to as accidental are those which he considers follow a probability distribution. He explains nonsymmetrical distributions as those arising when the accidental causes have unequal chances in different directions. It is in the study of these causes that most of Quetelet's statistical analyses are made.

First of all, to study a particular cause, its effects must somehow be isolated from those of the other causes. This is no trivial matter, since all that we can see is one effect for each data point observed. Quetelet devotes several sections to just this problem and gives several methods.

The study of accidental causes can be undertaken by looking at the distribution of the observations, perhaps with a histogram, and comparing it with known distributions. Quetelet was a great believer in the method of graphing data. Accidental causes can be eliminated by looking at means made up of many observations. He justifies this by Poisson's "Law of Great Numbers"[18]. Variable causes may be studied by comparing groups of observations which occur under the same circumstances:

> To discover variable causes, the most simple mode is to divide into groups or series the objects supposed to be under their influence. When these groups are formed in the same manner, and are in all respects comparable, they will be successively equal one to another, if the causes which have given them birth be constant. On the contrary, they will be unequal if the causes be variable. [19]

Variable causes may be eliminated by looking at these groups individually. This seems to be the first mention of an idea resembling "blocking" or "stratification". Periodic causes may also be eliminated or studied as one wishes:

> When the existence of a simple periodical cause is suspected, it is easy to study it, by comparing the different parts of the supposed period with another. Thus, if we wish to know whether the mortality is influenced by the period of the year, we must compare the results of the different months of the year. We shall find that the mortality is subject, at an interval of about six

> months, to a *maximum* and a *minimum*. In our climate the *maximum* occurs in January, and the *minimum* in July; between these two limits the other numbers increase and decrease with regularity.
>
> The influence of periodical causes may be easily eliminated by carrying the means which we compare throughout the whole period. Thus we had no difficulty in comparing the mortality of one year to that of different years, although the mortality varied considerably in the course of each. It is enough if the fluctuations are generally the same, which is in fact the case.[20]

Later on in "*Letters ...*" he gives an excellent example of how to conduct a statistical study of causation:

> In statistical researches, our object is generally to discover the causes which influence social facts, and to determine the degree of their energy. These appreciations, especially the latter, sometimes become impossible; and we must confine ourselves to the study of the causes, and their tendency.
>
> Society is not like a physical instrument which we arrange and disarrange at will for the purpose of studying it under all its forms, in all its variations, and at the most favourable time. Woe to him who shall attempt such experiments! He would have to complain of a country which every instant changed its laws, its customs, and its relations, to find by experiment the most suitable mode of existence. We cannot then, as in the greater part of the sciences of observation, equalize at will all the influential causes, save one, so as to study the effects and modes of action of this latter. We must often proceed by other ways: we must substitute analysis for synthesis, and commence by taking the phenomenon in its most general state.
>
> Let us remark, in the first place, that it is nearly always possible to disregard in the study of social phenomena, as in that of physical phenomena, the effects of *accidental* causes, by making the results depend on a sufficient number of observations. Thus, in studying the mortality of Belgium, I made the results depend on very large numbers, in order that we might consider the effects of accidental causes eliminated.
>
> It is equally possible to throw out the effects of periodical variable causes, by only comparing with one another results given by an entire period, or by the corresponding parts of a period. Thus, in studying the mortality of Belgium, I should establish comparisons between the results of successive years. In default of the results of a whole year, I should compare those of many successive winters or springs, because I suppose a constant ratio between the number given by one particular season and the whole year.
>
> If, instead of eliminating, we wish to study the effects of the period, we must operate in quite a different manner; it will be the partial numbers of the different divisions of the period that we must place before us.

> The study of <u>variable</u> <u>causes</u> <u>which</u> <u>are</u> <u>not</u> <u>periodical</u> is more difficult. The best thing to be done, in such a case, is to compare the result which is supposed to be influenced with many other similar results obtained for other years, so as first to judge whether there are any anomalies or sensible variations in these results. When a difference is found, we must discover whether it is periodical or fortuitous. We must next compare the result which we suppose to be influenced with the causes which may have made it vary. For example, I have shown in one of my preceding letters that the mortality in the years 1835 and 1836 was less than during the contiguous years. I was led to think that this diminution in the deaths was owing to the price of provisions; and to assure myself of this, I compared the returns of the deaths for the ten years from 1831 to 1840 with the prices of wheat and rye. I found that, in fact, the progress of these two statistical elements was nearly parallel. I concluded from this that the variable causes which produced a diminution of mortality was probably due to the fall in the prices of grain.
>
> When we wish to study <u>constant</u> causes, we must commence by eliminating the influence of accidental causes, and have regard to the variable causes, to clear also of their effects the results on which we work. It is for these reasons that the researches should rather be carried out on years which are complete, and not anomalous. [21]

When he cannot distinguish between two causes because they always occur one with the other, he calls them "confounded", the term which is still in use today to describe such situations.

In the above example, one may see several other statistical concepts which were due to Quetelet. The first is the idea of indirect measurement, in that it is not possible to measure causation directly. One must rely upon the observed effects to lead him back to the causes. This principle is in such wide use today that one might never think twice about it, but one has only to look at the 19th century reviews of Quetelet's work to see the shock that this idea produced. Meitzen, writing in 1891, reworded to modern terminology, calls this idea cute but dumb. Quetelet suggested measuring man's moral qualities by his actions and his mental qualities by their products. Today a statistician might measure intelligence with a series of IQ tests (the products of intelligence) but the statistician of 1835 would have insisted that intelligence cannot be measured and for trying to do so would have called today's statistician a "materialist", an insulting term of the day.

Another concept made use of in the passage is that of correlation. Quetelet sees that the mortality rate varies as the price of provisions. Walker claims that Quetelet was the first ever to use this concept, and Yule gives a formula that he calls "Quetelet's formula" for finding the correlation between two sets of random variables:

$$\phi = \frac{(AB)(B) - \frac{(A)}{(u)}}{\frac{(A)}{(u)}} \quad .^{22}$$

Quetelet further extends this idea of correlation to what one might call elementary regression in the next example, covering letters XXXII and XXXIII in "Letters ...". This example concerns the estimation of the blooming date of lilacs and is certainly the best example of modern statistical inference to be found for many years. In this example, he actually constructs a prediction equation for determining the exact date of blooming. The date is N in:

$$4,264 = \sum_{i=\text{first day of spring}}^{N} X_i^2$$

where X_i is the temperature on the ith day in degrees Celsius. The first day of spring is determined by the day the winter frosts end. When a late frost occurs, these frost days are not counted. His model in fact has a very good fit over the six-year period of investigation. He also determines all of the other causes that may be involved:

> I range in four principal classes the causes which may influence the blooming of plants.
>
> Geographical causes, such as the latitude, the longitude, and the altitude.
>
> Local causes, such as the nature of the soil, the exposure, and the quantity of light.
>
> Individual causes, such as the age and vigour of the plant.
>
> Meteorological causes, such as the temperature, the nature of the winds, the moisture of the atmosphere, the quantity of rain, and the state of the heavens. 23

He then tries to include other causes in his model in order to make it more general, but he meets with poor results. He has friends from all over the world gather information of the date of blooming of lilacs and send it to him. He notes their altitudes and latitudes. Comparing results, he estimates that there is a 4.1-day difference in blooming date per degree of latitude (X_{1i}) compared with that of Brussels (Z_1) and a 6-day difference per 100 yards in altitude (X_{2i}) compared with that of his own flower patch (Z_2). Letting X_0 be the blooming date of his flowers in Brussels and Y be the blooming date of all others, his prediction equation is:

$$Y = X_0 + 4.1(X_{1i}-Z_1) + 6(X_{2i}-Z_2) \quad .$$

He finds that this equation gives a poor fit of the data. We can assume by Quetelet's method of calculating the coefficients separately that he didn't know about, or it didn't occur to him to use, the method of Least Squares. But one thing that we certainly can assume was that Quetelet was far ahead of the other statisticians of his day. He gives his conclusion:

> I think it right to remark that the following consequences result from my mode of calculating blossoming.
>
> 1st. The line which may be drawn on the globe, through all the places where a plant may flower on the same day, is not necessarily parallel to the line which passes through the places where this plant flowered ten or twenty days earlier.

> 2nd. The <u>isanthesic</u> lines — that is, the lines of simultaneous flowering — have no general character, but vary at different periods of the year, and tend to approach the isothermal lines.
>
> 3rd. The time which elapses between two successive phases of the same plant may not be the same in different parts of the globe. If in England, for example, fifteen days separate the epoch of the putting forth the leaves from that of the flowering of the lilac, this interval would not be so long in Italy or Spain, where the temperature increases more rapidly. 24

One last contribution which Quetelet made to modern statistics is in the area of optimization. He makes brief mention of it in letter XLV of "<u>Letters ...</u>":

> Only to quote one example. There is a dependence between the number of travellers transported each day and the fares they have to pay; this dependence is such that the receipts augment or diminish according to the scale of fares. Everyone can conceive, in fact, that if the fares were too low, the number of travellers, although more considerable, would not be sufficient to pay the expenses of the enterprise; if, on the contrary, they were too high, the number of travellers would diminish, and the administration again would run the risk of a loss. There is then a <u>maximum</u> which can be obtained, and which can only be determined by the aid of good statistical documents. 25

He also includes in this letter several examples of optimization problems and several different but very modern examples of the use of statistics. It is certainly well worth reading.

Improvement of Data Collecting Techniques

Another of Quetelet's gifts was a sharp eye for determining good data from bad and a strong intuition for how to obtain good data. He realized the power of statistical methods and also that these methods were only as powerful as the data was accurate. He points out many errors that are committed by statisticians as well as others involved in the collection of data:

There are some stumbling blocks to be avoided, on which many statists (too little on their guard against the digressions of their imaginations) have fallen. The principal are the following:--

1st. Having preconceived ideas of the final results.

2nd. Neglecting the numbers which contradict the result which they wish to obtain.

3rd. Incompletely enumerating causes, and only attributing to one cause what belongs to a concourse of many.

4th. Comparing elements which are not comparable. 26

These oversights often lead to erroneous results, some of them obvious, some not. He gives an example of a friend who examines crime statistics and notices that in Belgium in 1820, 16 out of every 100 persons tried were acquitted, and in 1830 32 out of every 100 were acquitted, and concludes that soon there will be no criminals in Belgium. However, Quetelet points out that the jury system wasn't implemented until 1830, so these numbers are not comparable. His friend also made the mistake of "extrapolating to the limit", something else statisticians should not do. Sometimes this is done intentionally by people who wish to make a false point, especially in politics:

In politics especially statistics become a formidable arsenal from which the belligerent parties may alike take their arms. These arms may be accommodated to all the systems of attack and defence. Some figures, thrown with assurance into an argument, have sometimes served as a rampart against the most solid reasoning; but when they have been closely examined their weakness and nullity have been discovered. Those who allowed themselves to be frightened by such phantoms, instead of looking to themselves, prefer rather to accuse the science than to confess their blind credulity, or their inability to combat the perfidious arms that were opposed to them. 27

Documents should always be checked. One should prefer impartial data whenever there is a choice. He gives an example of how the U.S. government began distorting the facts at its birth:

During the war of independence, the United States carefully misrepresented the true number of their population: they exaggerated considerably the number of inhabitants in maritime cities,

> in order to put the enemy on the wrong scent. Assuredly, no good appreciation of the American population could be founded on the documents of this period. 28

His list of things to look out for include: small sample results, large accidental or periodic causes, unrepresentative samples and causes which are not obvious. He say that graphing the data can often show up irregularities in the data which could not otherwise be detected and also shows the relationships between variables. Quetelet gives his own procedure for dealing with "outliers" or oddball data points:

> ... If I find a number too small or two large, I check the calculations: if they are right, I check the sources from whence the original numbers were taken; if they also are exact, I seek whether the variation remarked is due to accidental causes: or whether it depends on constant ones.
>
> This study is singularly facilitated by diagrams. A simple line allows us to appreciate at a glance a succession of numbers which the most subtle mind would find it difficult to retain and compare. 29

He says that when one prepares a report for publication, one should always give the sample sizes, means, ranges, scales of precision, tell how the data were collected and give any irregularities encountered. One should also present the data in a clear form, i.e. tables or graphs, so that all relevant information can be perceived at a glance and not clutter up the report with irrelevant numbers.

When one is preparing a survey, one should ask only pertinent questions. Quetelet gives guidelines:

> 1st. Only ask such information as is absolutely necessary, and such as you are sure to obtain.
>
> 2nd. Avoid demands which may excite distrust and wound local interests or personal susceptibility, as well as those whose utility will not be sufficiently appreciated.
>
> 3rd. Be precise and clear, in order that the inquiries may be everywhere understood in the same manner, and that the answers

> may be comparable. Adopt for this purpose uniform schedules which may be filled up uniformly.
>
> 4th. Collect the documents in such a way that verification 30
> may be possible.

Above all, the statistician should know as much as possible about the area in which he is conducting his survey in order to be able to avoid illogical results and as many of the aforementioned errors as possible. This would be sound advice for a statistician of today.

Conclusion

Before 1830, statistics to some was the art of gathering data in order to better describe the state, and to others it was the study of populations. Probability theory was a branch of higher mathematics which had about as much relevance to the common man as any other branch of higher mathematics, which is to say very little. Between 1830 and 1850, Adolphe Quetelet succeeded in combining these branches of science into what is now known as statistics, and he truly deserves to be called the Father of Modern Statistics. He laid the foundations for modern sociology and showed that his new statistical science could be applied to meteorology, anthropology, astronomy, psychology and medicine. His letter XLIV in "Letters ...", entitled "On the use of statistics in the medical sciences", is really quite humorous as well as enlightening as to the ideas of the day and well worth reading. His well written books and numerous publications helped to spread his methodology far and wide. His contributions to the foundation of statistical societies and journals and his drive for uniformity in international statistics were directly responsible for the state of the science today. He is responsible for what is now the custom of always giving sample size and standard deviation with every mean.

Although some of his ideas on morality may seem foolish by today's standards, for the most part his work is still highly relevant and many modern statisticians could profit from reading "Letters ...", as well as be entertained by its marvelous style and often humorous examples. Adolphe Quetelet, statistician, astronomer, mathematician, physicist, sociologist, musician, poet and historian was certainly one of the great men of the 19th century.

Bibliography

1. Adams, William J. The Life and Times of the Central Limit Theorem, Kaidmon Publishing Co., New York, 1974.

2. Cournot, M. Exposition de la Theorie des Chances et des Probabilities, Librairie de L. Hachette, Paris, 1843.

3. David, F. N. Games, Gods and Gambling, Charles Griffin and Co., Ltd., London, 1962.

4. Federer, W. T. Statistics and Society, Marcel Dekker Inc., New York, 1973.

5. Gnedenko, B. V. The Theory of Probability, Chelsea Publishing Co., New York, 1962.

6. Hankin, Frank H. Adolphe Quetelet as statistician, Ph.D. Dissertation in Political Science, Columbia University, 1908.

7. Koren, John. The History of Statistics, Their Development and Progress in Many Countries, Macmillan Inc., New York, 1918.

8. King, Amy and Read, Cecil. Pathways to Probability, Holt, Rinehart, and Winston Inc., New York, 1963.

9. Meitzen, August. History, Theory, and Technique of Statistics, translated and with an introduction by Roland P. Faulkner, American Academy of Political and Social Science, Philadelphia, 1891.

10. Quetelet, Adolphe. A Treatise on Man and the Development of His Faculties, William and Robert Chambers, Edinburgh, 1842. A facsimile reproduction with an introduction by Solomon Diamond, Scholars' Facsimiles and Reprints, Gainesville, Fla., 1969.

11. Quetelet, Adolphe. Letters Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the Theory of Probability as Applied to the Moral and Political Sciences, translated and with an introduction by Olinthus Gregory Downes, Charles and Edwin Layton, London, 1849.

12. Seal, H. L. The historical development of the Gauss linear model, Biometrika 54, pp. 1-24, 1967.

13. Sleynin, O. B. On the early history of the law of large numbers, Biometrika 55, pp. 459-467, 1968.

14. Todhunter, I. A History of the Mathematical Theory of Probability, G.E. Steckert and Co., New York, 1931.

15. Walker, Helen M. Studies in the History of Statistical Method, Williams and Wilkins Co., Baltimore, 1929.

Notes

1. Meitzen, August. History, Theory and Technique of Statistics, translated and with an introduction by Roland P. Faulkner, American Academy of Political and Social Science, Philadelphia, 1891, p. 15.
2. Ibid., p. 22.
3. Hankin, Frank H. Adolphe Quetelet as statistician, Ph.D. Dissertation in Political Science, Columbia University, 1908, p. 46.
4. Ibid., p. 48.
5. Adams, William J. The Life and Times of the Central Limit Theorem, Kaidmon Publishing Co., New York, 1974, p. 32.
6. Ibid., p. 13.
7. Ibid., p. 22.
8. Ibid., p. 27.
9. Walker, Helen M. Studies in the History of Statistical Method, Williams and Wilkins Co., Baltimore, 1929, p. 18.
10. Adams, op. cit., p. 34.
11. Ibid., p. 45.
12. Hankin, op. cit., p. 12.
13. Quetelet, Adolphe. Letters Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the Theory of Probability as Applied to the Moral and Political Sciences, translated and with an introduction by O. G. Downes, Charles and Edwin Layton, London, 1849, pp. 92-93.
14. Ibid., p. 96.
15. Ibid., p. 278.
16. Ibid., p. 107.
17. Federer, W. T. Statistics and Society, Marcel Dekker Inc., New York, 1973, p. 23.
18. Quetelet, op. cit., p. 145.
19. Ibid., p. 134.
20. Ibid., p. 140.

21. *Ibid.*, pp. 147-148.

22. Walker, *op. cit.*, p. 130.

23. Quetelet, *op. cit.*, p. 158.

24. *Ibid.*, pp. 172-173.

25. *Ibid.*, p. 238.

26. *Ibid.*, p. 211.

27. *Ibid.*, p. 214.

28. *Ibid.*, p. 202.

29. *Ibid.*, p. 206.

30. *Ibid.*, p. 196.