Chapter Two
Historical Overview: The Emergence of Statistical Thinking

> *"Statistical literacy can be interpreted as meaning an ability to interact effectively in an uncertain (non-deterministic) environment"* (Hawkins, 1996, p. 2).

2.1 Introduction

Statistics and statistical thinking are major factors in everyday life and in many occupations. Media reports present, for example, opinion polls, social statistics and medical and financial data. Nearly every occupation deals with data, both quantitative and qualitative. Therefore there is a need to have statistically literate citizens and to create a way of reasoning that will produce an informed society.

A premise emerging from an ongoing debate about the nature of statistics is that statistical thinking is a fundamental independent intellectual method (Moore, 1990; Ullman, 1995). Falk and Konold (1992) hold the opinion that probabilistic thinking is an inherently new way of processing information as the world view shifts from a deterministic view of reality. This presumes that there is a deterministic view of reality which believes all observed phenomena have a cause and a non-deterministic view of reality which believes that randomness will explain certain phenomena. If a non-deterministic perspective is taken then the reaction is not to look for causes but to attribute the observed phenomena to randomness. Neither of these two polarised views captures the middle ground proposed by Joiner and Gaudard (1990, p. 35), albeit from a quality assurance perspective:

> *"Variation is not a new concept. What is new is the awareness of variation and how it affects everyday activities is infiltrating the workplace. . . . Knowledge of the theory of variation alters people's view of the world forever. It influences practically every aspect of how companies are managed."*

This variation perspective searches for causes in order to reduce the variation but recognises that some variation will be dealt with by attributing it to randomness. The focus is on how one acts and thinks in the presence of variation.

Thus it would appear that applied statistics depends on different reasoning processes to that encountered in pure mathematics (Begg, 1995; Hawkins, 1996). Herein lies a problem that statistics is often taught and treated like a mathematical discipline within the classroom. Many authors are stating that statistics cannot be taught like mathematics and that statistics should be moving away from mathematics back towards its roots as a scientific enquiry process (Cobb, 1991; Biehler, 1994a). Furthermore, with the advent of exploratory data analysis techniques there is a debate about the place of probability in data handling. Some people are of the opinion that probability should be removed from the teaching of introductory statistics while others believe that close links between probability and statistics should be developed in the teaching process (Shaughnessy, Garfield & Greer, 1996).

According to Porter (1986, p. 315) *"the intellectual character of statistics"* had been crystallised by 1900 and modern statisticians generally perceived *"the history of their field as beginning with Galton, if not Pearson."* Since statistics is a relatively new phenomenon, a brief historical overview up to 1900 will give some context and understanding on how statistics has been viewed and used, which may provide some insights into its present day views and uses. These insights may be helpful for the interpretation of the research data. Also, in order to understand and appreciate the nature of the debates on statistical thinking and on the teaching of statistics, the historical roots of statistics and the major conceptual barriers in thinking, that had to be overcome for the advancement of statistics as an independent discipline, should be considered.

2.2 Early History from the Renaissance Period

Although census data have been collected for millennia, the origin of statistics in its modern sense can be plausibly traced from the Renaissance period. Historically statistics is inextricably linked with probability, the roots of which can be traced to the solution of gambling or betting problems and to the handling of statistical data for mortality tables and insurance rates (Lightner, 1991).

One root of statistics is firmly placed in games of chance, which have been recorded as far back as 3500 B.C., and yet over some thousands of years no link was made with mathematics. It is theorised that this was because there was no concept of 'randomness', as it was generally believed that the outcomes of events were predetermined by some supernatural force. It was not until the Renaissance, when philosophical thinking began to change, that mathematicians turned their attention to chance in gambling games. A proto-typical Renaissance man Cardano (1501-1576) gambled daily and was the first to write a mathematical analysis of the probabilities of drawing aces out of a deck of cards and of throwing sevens with two dice. However the problem that is credited with giving rise to mathematical probability is the "problem of points". This problem dealt with the division of the stakes of a game of chance, between two equally skilled players, when the game was interrupted prior to its completion. Cardano unsuccessfully worked on the problem. De Mere, who made a precarious living at gambling, presented the problem to Pascal in 1654 and a solution was finally found in conjunction with Fermat. Thus through working on this problem, Pascal and Fermat, jointly laid the foundations for probability theory and ultimately statistical theory.

The other root of statistics is found in the analysis of quantitative data. Significant statistical investigations began only when insurance merchants needed probabilistic estimations of events. John Graunt in 1662 was the first person to draw statistical inferences from data. Through analysing data on the number of burials in various London church parishes he observed that: more male births occurred than female births; women tended to live longer than men; and the number of persons dying (except during epidemics) was fairly constant from year to year. Besides calculating insurance or annuity rates, the political arithmeticians, as they were known, were promoting the notion that state policy should be well informed by the use of data. They hoped that the use of data would consolidate state power and thus sideline the power of the church and nobility. The political arithmeticians in Britain, Germany and France had no reliable method of measuring population size as *"census data were non-existent"* (Porter, 1986, p. 20) and birth registers were incomplete, so they assumed that a prosperous community had an increasing population. Therefore cities almost always showed a declining population which was attributed to idleness and corruption. British and German authors moralised on the Catholic church and blamed the celibacy of priests for the depopulation of the popish lands. Sussmilch (1740) used parish records to estimate population numbers and wrote three volumes on population based on the biblical tenet 'be fruitful and multiply'. He believed that alcohol, gambling, prostitution, urban life, celibacy of priests and war were anathemas for increasing the population. He promulgated notions that the state should provide medical care as it would reduce the death rate, and that the distribution of property to all heirs would promote marriage and discourage emigration. He advocated that the state should expand its apparatus for the collection of population numbers and to use this information for creating state policies.

An analysis of the thinking, from the first root, reveals determinism in beliefs about games of chance, and from the second root, that commonly held community assumptions or deterministic beliefs could be used to create "data". Thus, behind the collection of data, there could be people who had reasons or personal beliefs for carrying out such an investigation. These two roots gave rise to the development of statistics into two separate parallel paths, with many linkages (Cheung, 1998), and to a fundamental conflict in defining probability. The gambling root gave rise to an abstract theoretical *mathematical definition* of probability whereas the other root in data gave rise to a practical *statistical definition* of probability based on actual experience (Green, 1982). It should be noted at this stage that these and other definitions of probability still cause much debate today but, for the purposes of this research, only the main arguments will be alluded to.

2.3 Early Attempts to Reconcile the Mathematical Definition and the Statistical Definition of Probability

Early works on probability were concerned with games of chance which provided simple and idealised situations to model probability. According to Laplace in 1812, in his publication *Analytic Theory of Probabilities*, the probability of an event is the number of ways it can occur, divided by the total number of things that can happen, assuming that all of the latter are equally likely. The a priori computation of chance for a game employing dice was straightforward, as was the chance of obtaining a red ball from an urn known to contain r red balls and b black balls. The a posteriori question of determining r and b based on observations of drawings from the urn proved to be more troublesome. In *Ars Conjectandi*, published posthumously in 1713, Jacob Bernoulli (1654-1705) attempted to reconcile the difference between the a priori mathematical probability and the a posteriori statistical probability. He attempted to give a formal mathematical treatment to the notion, that the greater the repeated experimental trials, the closer one is to obtaining the unknown proportion. This has come to be regarded as the first law of large numbers. In today's parlance, Bernoulli found that he needed 25,550 experiments to learn the proportions of r and b, with a margin of error of 2%, and that this would be correct more than 99.9% of the time. To go from a qualitative intuitive judgement to a quantitative one was a large step. De Moivre achieved a substantial simplification of Bernoulli's first attempt at the quantification of uncertainty by approximating the binomial probability to the error function, or in today’s terms, to the integral of the normal distribution.

Stigler (1986) believes that no-one applied or extended these ideas as they did not provide an answer to the fundamental question of inference for empirical situations. It is interesting to note that Bernoulli asked*: "what about problems such as disease, weather, games of skill, where the causes are hidden and the enumeration of equally likely cases impossible? In such situations it would be a sign of insanity to attempt to learn anything in this manner"* (cited in Stigler, 1986, p. 65). There would appear to be a conflict in relating the urn device problem to the real-world problem. A stumbling block towards developing these connections or abstracting common properties seems to be the lack of equiprobability in the real-world problem, and the notion that prediction is impossible when there is such a multitude of causes. Another possible reason is that the use of graphs as thinking tools was not employed extensively until the nineteenth century (confirmed by personal communication, Garry Tee, well-known mathematical historian, June, 1998).

2.4 The Transition towards Inferential Statistics

2.4.1 Breaking with the Gambling Interpretation of Probability

Probability and statistics entered a transitional period as mathematicians began to realise that *"many concepts from probability could not be separated from statistics, for statisticians must consider probabilistic models to infer properties from observed data"* (Lightner, 1991, p. 628).

Stigler (1986) posited that the chief conceptual step taken towards the application of probability to quantitative inference involved the inversion of the probability analyses of Bernoulli and De Moivre. He stated that there were two critical key ideas that encouraged Bayes' ground-breaking inference work in 1764. The first key idea was to not think in terms of games of chance, and the second key idea was from Simpson in 1755, who had a conceptual breakthrough in an astronomy problem. It is interesting to note that Bayes' work was conceived around a geometric diagram.

The astronomers' problem was that if 5 observers recorded 5 different times for the passage of a star past a crosshair in a telescope, how were the numbers to be reconciled? The astronomers' answer was to discard any number that appeared to be way out and then take the arithmetic mean. Simpson focussed on the errors made in the observations, on the differences between the recorded observation and the actual position of the body being observed. This was the critical step to open the door to an applicable quantification of uncertainty. He assumed a specific hypothesis for the distribution of errors. He was able to focus his attention on the mean error rather than the mean observation.

The methods developed by Gauss, Laplace and Legendre, such as the Normal distribution (known as the error distribution), the central limit theorem, the method of least squares, the combination of observations and the use of mathematical probability to assess uncertainty and make inferences, became commonplace in astronomy and geodesy but not in the social sciences. The reason for this was that astronomers could compare their observations with fact, and their predictions with the eventual reality, whereas the social scientists required massive amounts of empirical data before they were confident of using these methods. Also major conceptual barriers had to be overcome before this technology could spread to the social sciences.

2.4.2 The Statistical Interpretation of Probability

The main break with the mathematical or gambling interpretation of probability occurred in the 1840s when probability was explained in terms of the regularities produced by chance phenomena. In 1854 Boole insisted that a measure of probability was derived from the observation of repeated instances of the success and failure of events, and that the gambler's a priori reasoning was inapplicable. John Venn in 1888 gave probability its first full frequentist interpretation. He argued that gambling problems did not illustrate the theory of chance. He viewed probability as a science concerned with real data for purposes such as insurance and therefore it was based on experience. That is, historical relative frequencies with an underlying stability over time would provide probabilities of similar events in the future. He required that a probability of an event be determined by placing it in a series, and the probability value be applied to the series, not to the individual occurrence. For example, the probability of a male, between the ages of 20 and 25, being involved in a car accident would be predicted from that group's historical data. The predicted probability would then apply to that group of males, not each individual male. This was a fundamental shift in the interpretation of probability in that it was applied to a group.

2.4.3 Recognition of the Power of Statistics to Reveal

As the notion of probability for such purposes as insurance was being reinterpreted for statistics, there was also the notion present, that the improvement of the human condition could be achieved through knowledge gained from gathering and analysing data. Thus at the beginning of the nineteenth century there was a subtle shift in thinking for 'statistics' to be seen as a science of the state. The statists, as they were now known, conducted surveys of trade, industrial progress, labour, poverty, education, sanitation and crime. They considered that their task was to chart the course of economic and social evolution so that the confusion of politics could be replaced by an orderly reign of facts. There was a new set of attitudes, a new sense of power and dynamism in society after the French Revolution. Writers who wanted to instigate reform in orphanages, prisons and poor houses collected data which they misinterpreted and manipulated to promote their cause. For example, Taillandier, without any information on literacy rates of the population, found that 67% of prisoners were illiterate and therefore considered that he had strong proof that ignorance was the cause of crime. Hence he, along with others, promoted the idea that public education would lead to a reduction in crime. Statistical societies were set up. The French moral statists sought to control deviant behaviour by their 'numbers' and the British statistical society wrote a set of questions to investigate the effect of reforms on crime, education and public health.

However, people like Quetelet, a pioneer in social statistics, were reacting to the statistical reformers and started writing about the abuse of statistics to promote preconceived ideas. Politicians increasingly took cognisance of the 'facts' because they were silenced by the 'numbers' and without numbers legislation was considered to be ill-informed and haphazard. Alexandre Jonnes wrote that *"statistics does not have the power to act, but it* *does have the power to reveal"* (1856, cited in Porter, 1986, p. 29). *"The idea of using statistics for such a purpose - to analyse social conditions and the effectiveness of public policy - is commonplace today, but at that time it was not"* (Cohen, 1984, p. 102). The analysis of social data was hampered by inadequate data and a lack of statistical tools. Florence Nightingale (1820-1910) is regarded as a pioneer in this field as she sought to gather accurate hospital statistics through the development of a hospital statistical form in 1860. She also developed new graphical representations that, for example, dramatically revealed the extent to which deaths in the Crimea war had been preventable.

The interpretation of data is an aspect that is still debated vigorously today. The challenges about data interpretation, in terms of generalisability and other plausible explanations, are familiar arguments. However, the foundations of statistics as a science of data were being established through the frequentist interpretation of probability, and through the notion that knowledge about a situation could only be gained through data collection and analysis, not from personal opinion.

2.5 Breaking the Social Science Barrier

2.5.1 Expanding the Frequentist Interpretation of Statistical Probability

Quetelet from the 1820s onwards pursued a numerical social science of laws. He studied with Laplace and Fourier and argued that the true foundation of statistics had been established by mathematicians and astronomers. Quetelet applied the long run relative frequency idea of probability to social statistics. He looked at suicide and crime rates and was amazed to find large scale regularity. He interpreted this as proof that statistical laws could be applied to groups though not individuals. Through realising that general effects in society are produced by general causes and that chance could not influence events when considered collectively, he was able to recast Bernoulli's law of large numbers as a fundamental axiom of social physics. Porter (1986, p. 55) suggested that Quetelet's major contribution was in: *"persuading some illustrious successors of the advantage that could be gained in certain cases by turning attention away from the concrete causes of individual phenomena and concentrating instead on the statistical information presented by the larger whole."*

Writers and politicians were impressed with these statistical regularities. Lord Stanley, the leader of the Conservative party, claimed as an axiom that *"the moral and physical condition of the human race was governed by constant statistical laws"* (1856, cited in Porter, 1986, p. 57). Buckle, the author of the *History of Civilisation in Britain*, argued that these were general laws and gave examples of applications for society. For instance, in a given society a certain number of persons must put an end to their own lives, and if a particular individual did not yield to temptation, others would be impelled until the annual quota of crime had been reached.

Quetelet's revelations, through the application of the statistical frequentist interpretation of probability to social science data, led to much philosophising and debate about the ‘free will of man’. Thus a new way of processing information was instigating a new awareness of reality and a re-evaluation of determinism.

2.5.2 Moving from Data to a Statistical Model

There were deep conceptual barriers to overcome to adopt statistical methods to social data. In Quetelet's work on conviction rate data he could see the stability of the conviction rates in the categories and he wondered how he could measure the importance of the deviation from the average. He required a means of sorting the influences of different causes but lacked a statistical methodology. Since Quetelet was aware of the numerous causes and the year-to-year variations he was reluctant to aggregate counts. When Poisson analysed the conviction rate data his analysis differed in three respects: he fitted a probability model; he dealt directly with counts aggregated over the years; and he used Laplace's method of ratio estimation to measure the uncertainty of the overall rates he computed. He predicted using the model and he assumed juries remained stable over time.

Stigler (1986) believes that Poisson had this breakthrough because he was from an empirical science background and as an experimental scientist he was in control of his environment with relatively few variables whereas Quetelet had experience with census data. The central conceptual problem to extend statistics methodology from astronomical to social data was the isolation of social data into homogeneous categories. In order to apply the techniques of the theory of errors, the social scientists had to perceive the data within classes as amenable to analysis, and variation within a class as analogous to random fluctuations.

2.5.3 Interpreting the Statistical Model

In 1844 Quetelet announced that the astronomer's error law applied also to the distribution of human traits such as height. (His ‘average man’ was famous). He believed that nature designed according to a uniform pattern and then exhibited error. The analogy is that a statue is made and then 100 copies of it are made. The copies will exhibit error in height. He made a distinction between constant causes (average man) and perturbing causes (e.g. nutrition, climate). He lined up a table of the astronomical error function against the chest sizes of Scottish soldiers and these could be seen to be empirically the same. In the case of the heights of Frenchmen, who presented for conscription, he announced that there had been a fraud as the error curve fitted well except there was a surplus of heights just below 1.57 m, the height for exemption. Quetelet believed that all naturally occurring distributions of properly collected and sorted data followed a normal curve. His method lacked discriminatory power as too many data sets revealed evidence of normality. The method did not provide a key to evaluating and finding useful ways of classifying data for analysis and was not sensitive to more subtle types of inhomogeneities such as age or diet.

However, Quetelet's work created a climate of awareness that empirical observations could be modelled by theoretical distributions, and fostered a belief in the explanatory potential of statistical models. The striking fits convinced contemporaries that something like a 'central limit effect' was at work. It was proof that there was an averaging of random causes and it showed that nature could be counted to obey laws of probability. What Quetelet also did was to shift interest within probability from error to variation, and thus he provided the impetus for a reinterpretation of the error law as a law of genuine variation rather than mere error. Social science was promoting changes in statistical thinking.

2.6 The Recognition of Variation

In 1889 Galton was the first to use statistical methods of error analysis to analyse real variation in biology. For Galton the conformity of data to the normal curve was to be a ‘test’ for the appropriateness of classifying the data together in one group. The non-appearance of the curve was indicative that the data should not be treated together. His interest in hereditary, inspired by the findings of Darwin (his first cousin), led him to the assumption that if measurable quantities such as stature or exam scores followed the curve then the process could be inverted with respect to qualities that eluded direct measurement. Qualities such as talent or genius could be assigned a value on a statistical scale. For example 100 people's talent could be ordered and then assigned a score based on the normal curve. This notion was *“to become the most used (abused) method of scaling in psychological tests”* (Stigler, 1986, p. 271). The argument still remains weak today that a statistical scale is appropriate for talent because it is appropriate for stature. *“The method is so arbitrary that it is difficult to understand why it was ever taken seriously”* (Tee, 1991, p. 14). Such an issue, as to whether the measurement method does capture the relevant characteristic, would benefit from more debate in society today.

Galton's conundrum was that he could not connect the curve to the transmission of abilities from generation to generation. If the normal curve arose in each generation as the aggregate of a large number of factors operating independently what opportunity was there for a single factor, such as a parent, to have a measurable impact? Why did population variability not increase from year to year? The curve stood as a denial of the possibility of inheritance.

The breakthrough came with his pondering upon the size of pears on pear trees within a garden. His thinking was based on the question of what was influencing the size of the fruit. As an example, he considered the location (exposure to sunlight) of the pear trees and conjectured that the pear trees could be divided into three classes (tending to produce large fruit, moderate fruit and small fruit). The question then arose as to how a mixture of produce from all three sources could produce a normal curve. He theorised that the moderate phase occurred approximately twice as often as the extreme phases. He developed the quincunx as an analogy which "demonstrated" that the resulting mixture of normal distributions was itself normal (Fig. 2.1). Galton could conceive his data as a mixture of very different populations. In the view of Stigler (1986, p. 281), this was *"the single major breakthrough in statistics in the late nineteenth century."* From a mathematical perspective and definition this fact was known to such mathematicians as Laplace and De Moivre. It was Galton's conceptual use that was new and ingenious. It is interesting to note that in the use of the quincunx today the notion of the pins being an analogy for a physical factor has been lost. The context of the real situation has been stripped away to be usurped by a mathematical abstract approach.

Later Edgeworth provided a test to ascertain whether the walls from which two baskets of pears had been gathered had important differences. This necessitated determining the dispersion of the appropriate smaller curves, because that for the entire garden would be too large to serve as a basis for comparison. He estimated the variability internal to the subpopulations. He assessed the meaningfulness of the differences between the smaller populations. These tests were known to the astronomers who used them in a deterministic setting. Using Galton's conceptualisation of a large population being a mixture of smaller populations, Edgeworth extended these tests to measurements of quantities that were in effect randomly determined. These tests could be viewed as forerunners to the modern day t-tests and analysis of variance tests.

The conceptual breakthrough had been achieved and the classical theory of measurement errors had been rationalised for variation.

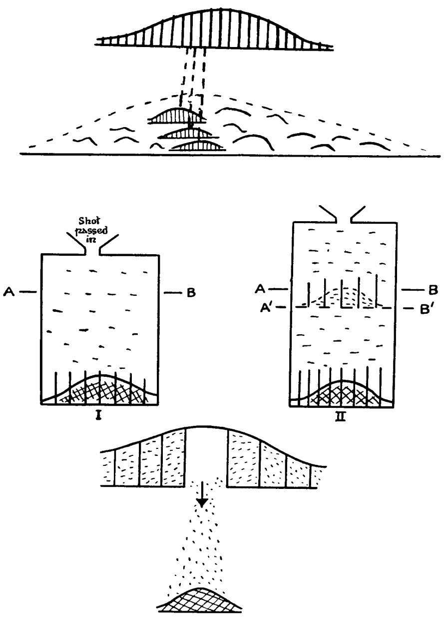

Figure 2.1 Galton's Two-stage Quincunx
(Drawings by Karl Pearson, based on Galton's writings, cited in Stigler, 1986, p. 279)

2.7 The Re-evaluation of Statistical Thinking

Debates about the use of statistics continued with the argument that statistical regularities proved nothing about the causes of things. There was recognition that, in statistics, the effects of constant causes were often masked and could be confused by other factors not considered, and that it was essential to sort out the effects of the perturbing causes before drawing conclusions. When Einstein declared that 'God did not play dice' he was reflecting the viewpoint of his times that scientific laws were based on causal assumptions and reflected a causal reality. The defence of human freedom inspired a wide ranging re-evaluation of statistical thought in the late nineteenth century. Statistics came to be seen not as a method of physical science applied to society but as a new scientific strategy. Chance was recognised as a fundamental aspect of the world in a way that it was not before. The acceptance of indeterminism constituted one of the noteworthy intellectual developments of the time. According to Porter (1986, p. 319) the evolvement of statistical thinking *"has been not just to bring out the chance character of certain individual phenomena, but to establish regularities and causal relationships that can be shown to prevail nonetheless."*

2.8 Summary

- Statistical thinking has its foundations in a frequentist interpretation of probability, and in the realisation that the gathering and analysing of data is needed for the acquisition of knowledge about a situation. That is, it is based in an empirical scientific method.

- Statistical thinking requires an ability to focus on the group propensity rather than individual cases in certain situations.

- Shifting attention away from gambling problems to other problem situations, and focussing on the error, were key steps for the development of inferential statistics.

- The type of data that is collected, and the way the data are manipulated and interpreted, are enmeshed in the prevailing attitudes, beliefs and prejudices of the society or the data analyser.

- During the nineteenth century the transition from descriptive statistics to inferential statistics came from a growing awareness, and from new conceptualisations, of how probabilistic models could be used to infer properties from observed data in a variety of domains of application.

- At the end of the nineteenth century the theory of measurement errors was reinterpreted as the theory of variation.

- The use of statistical ideas and models to explain human behaviour resulted in a new conceptualisation of the human condition as a group. The interplay between the statistical model and the real situation resulted in a new conceptual view of the world, causing a shift from a deterministic view of reality.

- Statistical reasoning evolved from new ways of perceiving and interpreting the mathematical models and definitions, and through the development of new tools for analysis.

This brief historical overview, and identification of some key factors in the development and nature of statistical thinking up to the beginning of this century, has provided me with some contextual background for interpreting and reviewing current statistics education research. A review of the literature, which I believed to be relevant to understanding the particular ways of thinking in the statistics discipline, is presented in the next chapter.