Presented to the Institute of Actuaries Students’ Society

on 7th January 1986

STATISTICAL METHODS FOR ACTUARIES

by

Patrick Carroll MA, FIA, FSS

Presented to a joint meeting with the
Royal Statistical Society
1. Introduction

1.1

Actuaries have taken on a hard task in statistical terms. Standard techniques that assume normality and independence of errors or employ additive linear models are not usually applicable to actuarial problems. Life Tables do not follow a Normal Distribution. Nor do the various risks which are the subject of General Insurance. The stock market shows a marked serial correlation which makes it inappropriate to assume independence of errors and special techniques of Time Series are considered more promising in this context.

2. Importance of Statistical Methods to Actuaries

2.1

The motto of the Institute of Actuaries is "Certum ex Incertis". This can be taken to imply a claim to expertise in applied probability and statistics. In 1984 a joint committee of the Institute and Faculty reviewed the Structure for Education and Training of the profession. In the present system of examinations two of the A subjects are regarded as statistics subjects and one as "actuarial statistics". The committee found that "There was difficulty within the Committee in deciding just what more advanced or specialised (statistical) topics were really relevant and should be included. However, the Committee agreed in general that the two statistics subjects should be restructured by the removal of some of the more elementary statistics and the inclusion of the more relevant more advanced topics ..."
2.2

Statistics is not just central to actuarial training and education but it is also the most contentious part of the long painful process. In his paper "Seven Years Hard" David Purchase said "The treatment of statistics in actuarial education has always been at the core of the argument." He also points to the introduction of General Insurance into the scheme of examinations (Subject B3) as having both lengthened the time taken to complete the examinations and having increased the emphasis on statistics. Whereas the traditional approach to Life Insurance and Pensions has been rather deterministic (and not unsuccessful) a statistical approach is more essential in General Insurance. We can all be pleased that General Insurance is now properly catered for. Actuaries are making progress with General Insurance. The problem is to find space in the professional syllabus for the relevant statistical methodology which is applicable to General Insurance and to Life Insurance and to Pensions and Investment.

3. Purpose of the Paper

3.1

The purpose of the paper is to describe various statistical methods that one might wish actuaries to know. It is assumed that some will merit inclusion in the new revised syllabus for what is now known as the A and B examinations. Other topics can be regarded as more suitable for an M.Sc. for actuaries. Statistical methods capable of application to actuarial work have developed extensively in recent years. The third edition of Risk Theory by Beard, Pentikainen & Pesonen is over 400 pages whereas the second edition was 200 pages. The level of mathematical and statistical methodology has also been somewhat enhanced. The American Casualty Actuaries also include in their list of reading a new
book titled Loss Distributions by Hogg & Klugman. This is very statistical and concentrates on fitting special parametric distributions to the sizes of claims in General Insurance.

3.2
A stochastic or statistical approach has also been taken more seriously by American Life Insurance actuaries. Norman Johnson who was once co-author of a text-book in statistics used in the Institute's reading list is co-author of a new text-book on survival analysis that is used in the reading list of the Society of Actuaries. Such methodology can be expected to replace now outdated topics like "Exposed-to-Risk Formulas" in actuarial statistics and mortality papers. The most profound change in the American reading list for those taking the Society examinations is represented by the new Life Contingencies text-book. In fact it is called Actuarial Mathematics but it goes a long way to superecede a deterministic approach with a stochastic approach. Actuaries should however be warned that not all developments in statistics are necessarily useful to them. Greater prominence given to Bayesian methods in a university course in statistics may lead to neglect of non-parametric methods needed by actuaries when no parametric distribution is known to fit their data. Likewise attention given to Markov processes may lead to the omission from the syllabus of the kind of specialised Time Series techniques that are applicable in actuarial work on investments. Time Series methods also apply in General Insurance when claims are known to follow cycles. Certain classes of General Insurance claims are seasonal. In other classes, cycles of several years duration are discernable.

3.3
Each of the major statistical topics needs to be considered on its merits so as to decide what can be included in the syllabus of the A subjects
that all actuarial students are required to take. There is also a need to allow those actuaries who wish to take a more specialised course in advanced statistical methods to do so while making some progress with the B examinations.

4. The Way Forward

4.1

Statistical Methods that might be useful to actuaries can be grouped under five main headings:

(I) Data Analysis

Ordinary methods of summarising and presenting statistics for display on a screen or on paper should not be neglected in a course of actuarial training. Pie-charts, histograms, stem-and-leaf diagrams, dot diagrams, bar charts etc. are not so trivial as to be left out of an actuarial syllabus. Likewise actuaries need to appreciate central measures and summary statistics that summarise data. In fact some sophisticated multivariate techniques such as Principal Components, Cluster Analysis and Correspondence Analysis could be included under the heading Data Analysis in that they do not use modelling or probability theory.

4.2

(II) Mainstream Statistics and Use of Transformations

Sometimes actuaries can use standard techniques that employ additive models and assume independent normal errors. English Life Tables have incorporated a Normal Curve to fit a special feature of the Table over a subsection of the age-range among Males as part of a multiparametric distribution to fit the whole Life Table. Motor insurance has successfully been investigated using a simple additive model. Whatever distribution is being fitted and whatever model is employed techniques
such as maximum likelihood may be considered practical for purposes of estimating parameters. Regression is widely appreciated among actuaries.

4.3
Simple transformations such as taking logs can make skew data distributions and multiplicative hazards more amenable to methods applicable to normal distributions with an additive model. A statistics course that is adapted to meet the needs of actuaries should include a discussion of transformations that can enable simpler models to be employed or standard assumptions to be made.

4.4
Various transforms as described in Risk Theory by Beard, Pentikainen and Pesonen have proved rather successful. These enable a useful approximation to Normality to be achieved for the distribution of aggregates of claims in General Insurance. Of course the Normal distribution is much easier to work with than Compound Poisson distributions.

4.5
(III) Non-parametric Methods
What actuaries should perhaps be most ashamed of not knowing are the non-parametric techniques that can be used to compare mortality in a group of lives with a standard table or with mortality in another group. Techniques for dealing with tied observations and censored data merit consideration.

4.6
Non-parametric methods do not however always produce answers. No parameters can mean no useful estimates. The scope of what can be done using purely non-parametric methods is limited. There may be a compensation here in that to include what is useful in the way of non-parametric methods would not be too heavy an addition to the syllabus.
(IV) Special Statistical Techniques for Actuarial Use

There is now a considerable literature on such distributions as the Inverse Normal and Pareto distributions. The mathematical theory associated with these distributions can be regarded as relevant to the extent that these distributions are applicable to the risks that are the subject of insurance. The Weibull distribution which is prominent in the literature on reliability theory may or may not prove to be of great value to actuaries. It has been featured in the new American Life Contingencies text-book called Actuarial Mathematics as well as in Loss Distributions. Generalised Extreme Value Distributions seem attractive as giving an overview of the sort of distributions that might be applicable to actuarial problems. As noted above special Time Series techniques such as Box-Jenkins ARIMA models which are not to be found in all university courses on statistics have found a prominent role in actuarial work on investment.

4.8

Some actuarial methodology has developed to be clearly beyond the scope of either the A or the B examinations. Such is the case with Premium Theory as described in a recent book by three Belgian authors.

4.9

(V) Non Mathematical Statistics. Use of Published Official Statistics, Data Archives and Market Research Surveys.

The Census provides much important demographic background for Actuarial work on Life Insurance and Pensions. Actuaries need some awareness of medical statistics and trends in health insurance for which government publications by OPCS are the main source.

4.10

Market Research data may not be made public in the same way as government
data. A large Insurance Office will at some stage wish to carry out its own Market Research survey. An ad hoc sample survey may typically employ interviews with questionnaires administered to perhaps 2,000 respondents.

4.11
Continuous market research has been successfully developed in many mass markets using a panel of regular informants. It is possible that the scope of continuous market research will be extended to embrace the insurance industry. Such a Longitudinal Approach can be a more powerful method of monitoring trends over time and can overcome the limitations of ad hoc questionnaire surveys. The OPCS Longitudinal Study is the demographic counterpart of continuous market research and it has been successful in overcoming the limitations of Death Certificates by linking with census records. Greater precision on occupational and social class measurements is thus achieved.

4.12
Some important practical techniques of actuarial statistics such as the “chain ladder” and other methods of estimating outstanding claim reserves are not to be found in the standard text books of mathematical statistics. Such techniques employing statistical ratios as coefficients embody regression principles and assumptions without explicit mathematical modelling.

5. Mathematical Prerequisites

5.1
Real Variable Analysis
The serious student of, say, the Pareto distribution will want to look at convergence and asymptotic properties. This will involve concepts of weak and strong convergence and use measure theoretical results. Such results of this kind as are useful can however be stated without proof and
it is not necessary to bring measure theory into the actuarial syllabus.

5.2

Complex Variable

Characteristic functions are mentioned in the third edition of Risk Theory and spectral densities are to the fore in Time Series work. Again the useful results can be taken without proof and no changes in the syllabus are called for. Rather one would like to raise the tone of mathematics in the reading list a little so that a good mathematician is not put off becoming an actuary. Must we continue with a statistics textbook like Wonnacott & Wonnacott that boasts an appendix explaining Least Squares without using calculus?

5.3

Linear Algebra

Here the institute needs to make a major policy decision. The decision that actuaries don't need to know matrices was presumably made before computers made matrix manipulation part of daily life. Once the use of vectors and matrices is admitted to the course of reading it is possible to describe in a clear and concise way a number of statistical techniques that actuaries might wish to know of.

6. Multivariate Methods

6.1

Multivariate methods continue to be a talking point in much of general insurance and in market research there are some clear illustrations of how such methods can be used. The advent of computers has both brought about the need for multivariate analysis by making the data available and also the means of doing the analysis. The actual work is largely matrix manipulation. The subject generally includes the following main topics:

6.2
The T-squared Test

Possible examples come easily to mind where measurements are made in pairs or in triplets so that a multivariate test is more appropriate than a univariate test. For blood pressures there are two measurements. For a Summary of Life New Business there is Nos of Policies, Sum Assured and Premiums. Convenient exam sized questions are no problem. The assumptions required of independence of errors and normality are similar to the t-test for which it is the multivariate equivalent. Starting from the t-test the T-squared test can be derived fairly concisely using the Union-Intersection Principle.

6.3 Regression

Whereas multiple regression is the most widely used research technique in fields related to actuarial studies the present A5 syllabus is confined to bivariate regression. Multiple regression allows for several explanatory variables on which a single response variable is linearly dependent whereas bivariate regression is restricted to a single explanatory variable. Once the use of matrix notation is admitted it is natural to introduce multiple regression. There is a large literature on applications of multiple regression and much of it is concerned with the problem of the choice of variables in multiple regression. The introduction of multiple regression into the syllabus is therefore needed to enable actuaries to embark on a major research project with the confidence of familiarity with the appropriate statistical methodology.

6.4 Principal Components

Principal components are perhaps the first of the distinctively multivariate methods for consideration when a classic multivariate problem of dimension reduction arises. A typical market research operation
might involve asking 2,000 respondents 100 questions. If these questions are in the form of a scale with quantifiable responses a 100 by 100 correlation matrix will emerge from the computer. To assist in identifying the important dimensions of consumers' attitudes multivariate methods can be called into play. Mathematicians are in the habit of finding eigenvalues given matrices and thus Principal Components, which can be found using the same computer calculations as eigenvalues, are first choice among mathematicians when faced with a multivariate problem of this kind. No modelling is involved. The first principal component is the linear combination of the measured responses that accounts for most of the variation (conveniently measured by the size of the eigenvalue). The second principal component is that which accounts for most variation subject to the constraint of orthogonality with the first. Third and subsequent components are found successively in a similar way. Principal Components tends to be regarded as a success when there are a few components that account for nearly all the variation and all are capable of a meaningful interpretation.

6.5
Factor Analysis

In the behavioural sciences it is unusual to stop at Principal Component Analysis. Rather it is common to employ Factor Analysis which employs a linear factor model and offers much more possibilities for attributing the observed responses on measured variables to a linear combination of factors which are variables not capable of direct measurement. Again the aim is dimension reduction so that there are fewer factors than variables measured and it is done in the hope of generating new ideas and hypotheses and gaining insights into consumer attitudes and buying behaviour. Important concepts such as "bargain consciousness", "economy mindedness", "quality consciousness" and "brand loyalty" may emerge either
as components in Principal Components or as factors in Factor Analysis. It is also usually possible for sceptics who don't believe either in market research or in multivariate methods to say that nothing has been discovered that is not otherwise obvious. But the techniques are much needed. How else can one make sense of a 100 by 100 correlation matrix?

6.6
Cluster Analysis and Principal Co-ordinate Analysis (Multidimensional Scaling)

Though the general aims of dimension reduction and making better sense of a large matrix are common to Cluster Analysis and other multivariate techniques it is rather exceptional among multivariate techniques being mathematical rather than statistical. The matrix used is a distance matrix rather than a correlation or covariance matrix. Statistical criteria for assessing the results do not apply and there is no agreed way of deciding between various methods of clustering. The technique is however worth mentioning to actuaries interested in market research and market planning. The Acorn system of using census data on housing and housing areas employs cluster analysis and it deserves serious consideration by those who wish to deploy direct marketing techniques more effectively. The aim is to identify and reach a well defined target population that can be sold a particular product more efficiently.

6.7
Correspondence Analysis

The latest fashion in Multivariate Analysis newly imported from France is called Correspondence Analysis. It is a technique which has an affinity with Principal Components and also with Cluster Analysis. It is especially useful for survey data as it can be used with categorical data and the results can be displayed graphically in a chart with an apparent meaning. Rows of a matrix (after centering) are plotted as points,
Distances such as a chi-squared measure are used for this purpose. It is a great advantage to show results in what seems to be a clear picture. So it can be expected that Correspondence Analysis will be widely used. But British statisticians are uneasy at the distributional behaviour of the points representing the rows in the matrix. Two rows with the same proportions come out as the same points. There is a scaling problem. It may be safer to use the technique as an adjunct to other methods of multivariate analysis rather than as an answer in itself.

6.8
Canonical Variate Analysis
When there are two sets of measured variables rather than one, Canonical Variance Analysis might be applied. It corresponds to using Principal Components on one set of variables. The first pair of canonical variates consists of the linear combination of the first set of variables most correlated with a linear combination of the second set of variables. The second pair is similarly defined subject to a constraint of orthogonality with the first. If a market research interviewer asks 50 questions about Life Insurance and 50 questions about Pensions, one might seek to use canonical variate analysis to assess how attitudes to one relate to attitudes to the other. Unlike regression where one set of explanatory variables explains another set called the response variables, in canonical variate analysis both sets of variables have equal status.

6.9
Discrimant Function Analysis
Here the aim is to decide to what population a new observed multivariate object should be allocated. Is a newly discovered fossil bone that of an ape or of a man? The statistical theory of discriminant function analysis developed by the great Sir Ronald Fisher and others is a good thing to have in a university course on multivariate analysis. One
suspects however that there is not the space in the professional syllabus nor do actuarial students have the time for discriminant function analysis.

7. Survival Analysis

7.1

Concepts familiar from actuarial studies of mortality reappear in a more probabilistic guise in Survival Analysis, which is also sometimes called The Analysis of Failure Times. What corresponds to the force of mortality is the instantaneous hazard rate \( \lambda(t) \). The proportion surviving is measured by the survival distribution function. The question of "Exposed-to-Risk" becomes a discussion on "Censoring". The usual form of censoring considered is sometimes called "right censoring" and it happens when an individual is lost to observation before he or she dies. A death before observation started to be made would presumably be called "left censoring". The more sophisticated literature on survival analysis is concerned with finer distinctions between different kinds of censoring. Type I censoring has been defined to refer to a situation where all individuals are observed for a fixed time. Type II censoring might refer to a scheme of observation until a predetermined number of deaths is reached. Random Censoring refers to individuals being removed from observation at randomly distributed times.

7.2

The principles of Survival Analysis are entirely consistent with traditional actuarial methods. The subject matter is certainly familiar to actuaries. What makes the subject seem strange is the use of the exact date of death and the non-use of a rate interval which is usually a year in actuarial work along traditional lines, e.g. calendar year, policy year, life year. The advantages of such an approach,
which is practical now that computers can count to an exact day without difficulty, are a better use of information and the greater potential for investigating the effect of explanatory variables on survival data. Whereas a standard life table requires an enormous amount of data for its construction it is possible to get useful results in Survival Analysis with a few hundred or even a few dozen lives and deaths.

7.3 Kaplan-Meier Product Limit Estimation

This is a simple technique for constructing the survival distribution function given the survival times and times to censoring of the lives under observation. The graph is a step function that starts at 1 and diminishes to zero (unless the longest survivor is censored before he or she dies). At each death the graph takes a step down. At a time of censoring there is no step. Rather the step at the time of the next death is increased. Actuaries will feel at home with the debate as to whether such step functions should really be smoothed or graduated. An unexpected attraction for statisticians is that the Kaplan-Meier method satisfies maximum likelihood criteria even though it is non-parametric. Even better news for actuarial students, if they have to learn it, is that the proof which considers a space of functions is not difficult.

7.4 Non-Parametric Tests for Survival Data

The explosion of activity among medical statisticians using new methodology has thrown much more light on the non-parametric tests applicable to mortality data. The log-rank test and the related Mantel-Haenzel test are prominent in the literature but also little known to actuaries. Modifications to the Mann-Whitney and Wilcoxon tests to allow for tied observations and censoring have received much attention. The Kolmogorov-Smirnov measure D is attractive as it allows for the
cumulative nature of mortality risks. All these techniques can be made easily accessible to actuarial students and provide actuarial examiners with suitable exam questions.

7.5

Parametric Approaches to Mortality: Distributions Applicable to Life Tables.

A feature of recent work on Survival Analysis is the stratification of observed data so that distributions which are simple and tractable are fitted successfully to sub-groups of observations though the whole experience cannot be well described by such distributions. Thus the Exponential and Weibull distributions are prominent in the literature though they don't really fit a whole Life Table. Also of course Gompertz and Makeham curves familiar to actuaries, can fit well cancer and heart disease statistics and deaths at more advanced ages in general. These distributions are not however popular with epidemiologists using Survival Analysis as censored data are not easily dealt with in the fitting of Gompertz or Makeham distributions.

7.6

Cox's Model

When Cox presented his 1972 paper on Regression Models and Life Tables it was enthusiastically welcomed by medical statisticians. As a successful synthesis of Life Table techniques which are central to the corpus of actuarial techniques and Regression, the "cornerstone of statistics" (Cox), it has now become an intellectual necessity for a good course in actuarial statistics. Cox's model has also become a big industry which actuaries can't ignore. The recent book by Norman Johnson and his wife was criticised by one reviewer for only having 20 pages on Cox's model. Actuaries in Life Reinsurance need to assess the results of medical statistics to prepare underwriting manuals for sub-standard underwriting.
So actuaries cannot afford to lose touch with the methodology of medical statistics. There is also some possibility of using Cox's model for investigating say persistency data in Life Insurance. Why do some policies lapse more than others? Perhaps an analysis using Cox's model could help? Tied observations are a considerable nuisance in the application of Cox's model and early lapses can be expected to fall on exact monthly intervals as most policies are the subject of monthly premiums. So persistency data does present a challenge. Also the shape of the mortality curve when heart disease or cancer is diagnosed is much different from the shape of the survival curve for newly issued life policies. The exponential term in Cox's model reflecting the influence of the explanatory variables produces a steeply increasing hazard rate when the parameters \( \beta \) are positive. However it might be possible to make appropriate adaptations to the model and the base line hazards using the kind of distributions that allow censoring to be accommodated while also fitting the shape of early lapses.

7.7

Cox's model is simple and concise to state (with vector notation)

\[
\lambda(x; t) = \lambda_0(t)e^{x'\beta}
\]

It is expressed in terms of hazard functions \( \lambda \).

The vector of unknown parameters used as coefficients of the explanatory variables is \( \beta \). The vector of observed covariates (e.g. age, sex, blood pressure, etc.) is \( x \). The quantity \( \lambda_0(t) \) is known as the base line hazard function.

7.8

The great merits of Cox's model are that the exponential terms lead to a relatively tractable log-likelihood and the model has been found very
suitable for both cancer data and heart disease data and other medical applications such as studies of transplants. The general approach to fitting the model is to regard the terms involving the base-line hazard function $\lambda_0(t)$ as nuisance factors in the first instance so as to estimate the parameters $\beta$. Then with $\hat{\beta}$ found, estimation of the base line hazard can be better effected. To eliminate the terms involving $\lambda_0(t)$ it sometimes seems necessary to assume a tractable parametric form for the base-line hazard, e.g. exponential or Weibull. So Cox's model is not as non-parametric as it looks and it is not simply a generalisation of Kaplan-Meier Product Limit estimation. A fully non parametric approach to the base line hazard function is also being explored.

7.9
Theoreticians have felt uncomfortable with the crude process of dropping off or ignoring part of the likelihood. It was pointed out that what Cox initially called a "Conditional Likelihood" was misnamed and the process of ignoring the nuisance terms involving the base-line hazard function is now described as a method of "Partial Likelihood".

7.10
A more satisfactory rationale for using maximum likelihood methods on such a likelihood function was developed using the concepts of marginal likelihood and order statistics. A rationale of this kind resembles that for using the log-rank test. The log-rank test does not involve logarithms but it does involve the idea of ranking or ordering.

7.11
Marginal likelihood does not however displace Partial Likelihood as being a better justification for all the use made of Cox's model. Time dependent covariates are not susceptible to ranking and ordering of this kind and so Marginal Likelihood arguments do not apply.
The basic method of maximum likelihood will no doubt be worth including in a revised syllabus for the A examinations together with a description of Cox's model. But there may not be space in the syllabus for the refined discussion of Partial Likelihood, Marginal Likelihood etc., nor for the practical use of Cox's model which is quite demanding as a computational exercise.

The subject has developed to provide a lot of good material for an M.Sc.
8. Time Series

8.1
ARIMA or Autoregressive Integrated Moving Average Processes developed by Box and Jenkins have proved capable of fitting to successive observations of stock market indices. The results can be used on prospective studies to investigate the effects of maturity guarantees on unit-linked life policies. Much use of forward projections is also made for preparing illustrations as marketing aids. A large and growing proportion of all life insurance policies are unit-linked and these policies are being sold as investments. The straightforward and objective process of curve fitting in the Box-Jenkins manner has commended itself to actuaries. There is in fact some resemblance between some Box-Jenkins formulae and some of the formulae traditionally used by actuaries for purposes of graduation (e.g. Spencer's 21 term formula).

8.2
Any curve fitting method used for forecasting is data dependent. If the last 50 years data say is used to fit the parameters in a Box-Jenkins model it is implicitly being assumed that the past 50 years' history of the observed process is a useful guide to what will happen in future years. In fact the last 50 years have seen a major change in thinking on the part of investors represented by the appearance of what is still called the "Reverse Yield Gap" around 1959. Since 1959 the gross dividend yield on a typical ordinary share has been less than the gross income yield on a typical British Government irredeemable stock. The reasoning in the minds of investors is that over the medium and long term the dividends received on the ordinary share will, it is hoped, show an upward trend. One can hope that this future growth will more than compensate for the lower initial yield. Prior to 1959 the higher risk of ordinary shares was regarded as a far more important factor than the
possibility of future growth and hence the initial yield on ordinary shares was typically higher than that on irredeemable gilts. One wonders what event in the future is anticipated corresponding to the appearance around 1959 of the Reverse Yield Gap by those who use the last 50 years to fit Box-Jenkins models to an index of ordinary shares.

8.3
Both the theory and the computing associated with Box-Jenkins methods are highly developed. Again it can be regarded as providing good material for an M.Sc. course. A one day course at The City University some years ago on this subject was popular and was rerun several times. Some introduction to the subject can be incorporated in a revised syllabus for subject A5 and there is already some mention of Time Series in the syllabus. It is interesting that the models found most satisfactory for investment applications incorporate a random walk component. Random Walks may thus earn a place on the professional syllabus which would also please the Markov process enthusiasts.

Matrices are quite prominent in Time Series literature.

8.4
More recent literature on Risk Theory points out the relevance of Time Series methods to the study of aggregates of claims in General Insurance. However little research has been done. This seems a promising area for new research initiatives. Trends and cycles are much discussed in relation to General Insurance and the subject seems a natural field of application for Time Series methods.

9. Generalised Linear Models - GLIM

9.1
Linear Model for Multiple Regression (and Analysis of Variance). The typical model for regression is of the form $y = X\beta + \xi$
where $y$ is the vector of observed response variables

$X$ is the known matrix of observed explanatory variables

$\beta$ is the vector of unknown parameters to be estimated

$\epsilon$ is the vector of random errors assumed to independent, of zero mean and with common variance.

The normal equations for estimating the unknown parameters follow using the method of Least Squares i.e. on minimizing the sum of squared errors $\epsilon'\epsilon$ with respect to the components of $\beta$.

The Normal Equations are $X'X\hat{\beta} = X'y$.

9.2

Thus far it is not necessary to assume that errors are normally distributed but to justify the Normal Equations by the method of maximum likelihood and to apply standard significance tests and methods of constructing confidence intervals it is usual to assume that errors are normal in the context of regression.

9.3

Linear Model for Factor Analysis

The model for Factor Analysis is of a similar form to that for Regression with the major differences that the factors are not directly observed and errors include variances specific to each factor. Again the use of correlations and methods such as maximum likelihood for fitting the parameters in the model require the assumption that observed variables are normally distributed.

9.4

Generalised Linear Models - GLIM

A major advance in statistical methodology has been the development of Generalised Linear Models (or Log-Linear Models to be more precise) applicable to integer valued random variables where the errors can be assumed Poisson or distributed other than normally.
computer package GLIM incorporates many facilities for the fitting of such models and displaying the results. The method of fitting is maximum likelihood by way of iteratively reweighted least squares. The measure of goodness of fit shown as "SCALED DEVIANCE" is a Welch type statistic being \(-2 \log (\text{likelihood ratio of observed and fitted likelihoods})\). This approaches a chi-squared.

9.5

GLIM is widely used to investigate contingency tables and it offers the possibility of assessing the relative size of influence of different explanatory variables. It is usual to assume Poisson errors and it is encouraging that the results of using GLIM seem fairly robust against departures from the Poisson in the error distribution. GLIM can be used with Negative Binomial errors and the results when such an error distribution is specified are little different from the results with Poisson errors.

9.6

There is considerable potential for using GLIM for purposes of analysis of General Insurance data. In General Insurance it is common for the number of claims to be known at an earlier date than the size of claims is known. It is important to make good use of such claim frequency data to obtain an early warning of possible deterioration in the portfolio of the relevant class of business. The number and type of claim can be regarded as a contingency table of integers suitable for GLIM. An example of an M.Sc. dissertation on Shipping data analysed using GLIM is provided by that of Miss Dina Shah at The City University. A comparable example also appears in the book by Nelder & McCullough.

9.7

Cox's model is not exactly linear or log-linear as it stands but it is possible to use GLIM to fit Cox's model when the base line hazards are in
a suitable form (Exponential, Weibull etc.) by using various tricks described by Clayton and others in a JRSS paper. With these distributions censoring can be dealt with.

9.8

GLIM can be used to carry out Analysis of Variance. The 'SCALED DEVIANCE' provides the sums of squares for an Analysis of Variance Table with appropriate degrees of freedom also shown. But GLIM is not designed for Multivariate Analysis. The GLIM program package is 7 or 8 thousand lines of Fortran so there is not much space left on the computer to do the matrix manipulation required for multivariate analysis. Using GLIM is rather like using an interactive language one level higher than Fortran. Keyboard signs such as % and £ are used by the GLIM system to execute Fortran programs. It does not seem appropriate to introduce GLIM or indeed any particular programming language or program package into the actuarial syllabus. There is no agreed way of presenting results from using GLIM. Some leading GLIM practitioners stress the importance of interpreting the parameters fitted. Others say it is unrealistic to try to interpret so many parameters. Overparameterisation is a temptation when using GLIM so as to achieve a good fit as measured by the 'SCALED DEVIANCE'.

9.9

The use of GLIM can however be recommended as part of an M.Sc. dissertation. The ideas which are prominent in the use of GLIM can be introduced in the syllabus of the Institute. One would like to feature in a revised syllabus for A5 such topics as the use of log transformations, least squares, maximum likelihood and the likelihood ratio principle. This would not unduly enlarge the existing syllabus. Poisson distributions are also familiar to actuaries. If the Analysis of Variance is retained in the A subject syllabus, consideration might
also be given to including the sort of models that can be fitted using GLIM. "Crossed" and "Nested" models can be used to illustrate both the use of GLIM and the Analysis of Variance.

9.10

Motor Insurance has been the subject of much research and is now quite well understood. Some quite elaborate models allowing for the many risk factors have been put forward. Problems of space on the machine, when GLIM is used with a large data file to fit models with many parameters, become acute. The subject is reviewed in a recent JIA paper by Coutts. The data considered by Baxter, Coutts and Ross is also considered by Nelder and McCullagh in their book.

9.11

An example of GLIM being used to analyse Life Insurance persistency data is provided by Haberman and Renshaw. The analysis of data from a Faculty of Actuaries source to which seven Scottish Life Offices contributed was quite successful. A simple model with additive mean effects in the systematic component without interactions was found to fit satisfactorily. Analysis of Variance showed the relative main effects of Age at entry, Duration, Office and Policy Type.

10. Bayesian Methods and Decision Theory

10.1

Bayes' theorem has long had an honourable place in the actuarial syllabus. With the appearance of Lindley's book in the 1960's there was much excitement among statisticians that an alternative approach to developing the standard results used in statistical methodology was possible and the Bayesian movement is still influential in academic statistics. The Bayesian approach is closely akin to the decision theoretic approach and in a university degree course the two are often combined as a single option in a final year. Important decisions in insurance do not however
permit easily a statistical solution. Whether to enter or withdraw from the market for a particular class of insurance may be decided on commercial grounds in relation to the connexions of the office and it may not be appropriate to state explicitly prior beliefs or utility functions. On the other hand there are in insurance some realistic null hypotheses worth considering and some important problems of estimation that can be tackled in a statistical and scientific manner. Is a given premium or reserve adequate? Does a particular experience accord with a standard table? What is the best estimate of a particular net premium? There is a great deal in the way of estimation and hypothesis testing that actuaries might wish to know. So it is tempting to conclude that no change is needed in the professional syllabus, as regards Bayesian methods and decision theory. Bayes’ theorem can stay in the A2 syllabus and a question on decision theory can be included in A5. It would however be preferable to update the syllabus at least by illustrating the use of Bayes’ theorem on continuous distributions. The Bayesian approach is taken seriously by Hogg & Klugman in the new book on Loss Distributions. Actuaries are aware of the advantages of checking results by using a different method and the Bayesian approach can either produce the same answer as the classical or frequency theory approach or produce a somewhat different answer. When a Bayesian approach is tractable it can provide a valuable check on the estimate of a parameter got by classical methods. In Factor Analysis one standard method of estimating factor scores results from a Bayesian argument. It is often more natural and advantageous from a teaching point of view to regard probability as a degree of belief. The idea of a confidence interval is less artificial from a Bayesian perspective. The great limitation of Bayesian methods is the requirement for explicitly parameterized prior distributions. In Bayes’ theorem: Posterior Prob. = Prior Prob. X Likelihood, the form of
the likelihood is determined by the nature of the data so the prior distinction must be in a form that combines with the likelihood to produce a tractable posterior probability. There is some uneasiness about the improper prior densities required when the likelihood takes a normal form (and some marginalisation paradoxes have been discovered in rather special multivariate situations). There is reason to doubt if the Bayesian approach is viable at all when the normal distribution does not apply as happens with most insurance data. So actuaries can't expect a great deal of practical help from the Bayesian school. It is however worth noting that the emphasis on likelihood methods that can be expected in a revised actuarial statistics syllabus is not really objectionable from a Bayesian point of view as the likelihood plays a central part in Bayesian analysis.

10.2

The Bayesian approach to Experience Rating and Credibility Theory in General Insurance is not unpromising. Jewell has shown that the incomplete gamma function can be used in conjunction with the mixed Poisson and other distribution mixtures to produce a tractable posterior density giving simple results. In this way a formula for experience rating that has a long history can be justified in Bayesian terms. The account of Bayesian Experience Rating in Risk Theory by Beard, Pentikainen & Pesonen follows what Jewell has done.

10.3

Also promising is the Bayesian approach to Time Series. A purely objective Time Series model such as is derived using Box-Jenkins methods cannot easily allow for known changes in the stock market causing sudden discontinuities in the level of prices for stocks. The government might for example announce major tax changes increasing or reducing the value of ordinary shares or gilts. It is then desirable to input an estimate of
what effect this will have into any model used for prediction purposes without delaying to collect say 20 observations to use for estimation of parameters. i.e. a subjective method is sometimes preferable to a purely objective method. Bayesian methods which are not yet as developed as Box-Jenkins methods may in the future attract more attention from actuaries. It would be nice to make available to actuaries an advanced course in Bayesian Statistics and related topics such as Decision Theory, Credibility Theory and Experience Rating. It would not be so nice to inflict the subject on all actuarial students taking the A and B exams as the demands made by the examinations are already perhaps too great. Some advantages can be claimed for the Bayesian approach to Statistics but it is nearly always more difficult mathematically than the classical objective approach using frequency theory of probability.

11. Special Distributions and Risk Theory Applicable to General Insurance

11.1

Teivo Pentikainen has said that he always uses empirical distributions for fitting or representing the size of claims for purposes of modelling and simulation to investigate solvency. It is important to avoid the biased approach of "have distribution-will-fit-it". On the other hand as Hogg & Klugman stress in their book "Loss Distributions" the fitting of a parametric distribution can give valuable insight into and understanding of a risk process. In certain classes of general insurance the smaller claims will be settled relatively soon while the larger claims can take years to settle. Once an appropriate distribution is known to fit the sizes of claims, information on the small claims can be used to predict the sizes and frequencies of the larger claims with advantages for the calculation of reserves and premiums. Real life data on claims contain
zeros which the fitting of a parametric distribution can usefully smooth out. There are theoretical attractions in using the Pareto and Generalised Pareto distribution for classes of insurance where exceptionally large claims or catastrophes are to be anticipated. The convergence properties of the Pareto seem appropriate in this context. Teugels has reviewed the extensive literature on this subject of generalised Pareto distributions. Other distributions listed as important by Hogg & Klugman include the Weibull, Beta, Gamma, Burr as well as familiar distributions such as chi-squared and exponential distributions. They also note that there are no conjugate distributions for most of these as required for Bayesian analysis. The fitting of special distributions and such important practical aspects as using truncated and mixed distributions is something to consider if space will permit to be included in actuarial syllabuses.

11.2

Risk Theory

In Risk Theory the emphasis is not so much on the distribution of claim size as on the frequency and multiplicity of claims. So the process is an integer valued one and Poisson and Mixed Poisson and Compound Poisson processes have become familiar in this context. The question of solvency is always relevant to actuarial thinking. For teaching purposes computer simulation is a rather crude tool from a statistical point of view. It seems unsuitable to include this in an actuarial syllabus though some might regard it as a suitable exercise for an M.Sc. dissertation.

11.3

Stop Loss Insurance and Reinsurance

Some leading reinsurance companies won’t write Stop-Loss contracts. But Stop-Loss is a good talking point in reinsurance. It is also a source
of good illustrations of Risk Theory and the stochastic approach to Life Insurance. So one would like to bring in Stop Loss Insurance to the syllabus for educational reasons. Reinsurance can provide useful exam questions. More seriously it is not possible to provide a comprehensive treatment of solvency without consideration of reinsurance. Stop Loss in relation to Group Life schemes would be a good addition to the syllabus for specialist Pension actuaries.

11.4 Premium Theory

This subject has developed a great deal as a mathematical subject that has become rather theoretical. The book by Goovaerts, de Vylder & Haerdonck was reviewed in a recent edition of J.I.A.

12. Bibliographic Notes

12.1 Life Contingencies

The new American textbook ACTUARIAL MATHEMATICS by Bowers, Gerber, Hickham, Jones and Nesbitt incorporates a radical change to a stochastic rather than a deterministic perspective. The reasons for such a profound change in the central textbook are twofold: advances in computers and advances in statistical methodology. Not until now have either been taken into account in a Life Contingencies textbook. In the new book however, pensions are really still treated in a deterministic way. There remains a lack of good examples to illustrate the use of a statistical approach to Life Table functions. Also the emphasis on means and standard deviations is not altogether satisfactory when Life Tables and the associated Life Assurance and Annuity functions do not follow a Normal distribution. The new textbook is rather long and does not correspond to just one exam in the scheme of the Society’s
examinations. Of the traditional topics that have been omitted actuaries may be sorry to have lost the treatment of stationary populations which is a deterministic way of modelling in demography but central to how actuaries think. In fact a deterministic model is always simpler and a useful check on a stochastic model. 12.2

The current British textbook Life Contingencies by Neill has been found a satisfactory treatment of the actuary’s traditional wisdom. Some of the approximate methods of valuation that computers have made redundant that appeared in the previous textbook by Hooker and Longley-Cook were omitted by Neill. But Neill still has expressions such as $1 + \frac{1}{2}i$ when all actuaries and actuarial students now have calculators that can take square roots at the touch of a button. Where the content of Neill seems most out of date is in the treatment of pensions. The refund of contributions is out of date. So are Children’s and Orphans Pensions that are unrelated to salary. One would like to include Transfer Values in a new treatment of Pensions and also bring in a new chapter on index-linked Life Insurance which accounts for a large and growing proportion of all new life insurance business. If Life Contingencies is to remain a single exam it might be better to retain the deterministic approach for this subject and introduce a statistical or stochastic approach in the examination on actuarial statistics. Profit testing which is rather deterministic and is something like a computer simulation could take its place alongside the Analysis of Surplus in a Final or B exam. Group Life can be treated simply and is so important as a class of insurance that one wonders how it can have been omitted for so long.

12.3

Multivariate Methods

The book to recommend is that by Mardia, Kent & Bibby which provides a good compromise between theory and practice making intelligent use of
matrices. An elegant mathematical treatment is provided by Dempster with admirable use of both algebra and geometry. A Bayesian approach is represented by S. James Press. The book by Morrison has some good examples and describes very well the application of multivariate methods to the behavioural sciences.

12.4 Survival Analysis

Several books have appeared in recent years. An obvious first choice is that by Cox & Oakes as it is not expensive and Cox initiated the main theoretical development of the subject. Also to be taken seriously is the book by Regina C. Elandt-Johnson and N.L. Johnson which gives a lengthy and thorough treatment of the relevant distribution theory and the fitting and testing of (distributions) models. It is more of a work of reference than a suitable textbook for actuarial students but it is used in the reading list of the Society of Actuaries. There is talk of a new American textbook better adapted to the needs of actuaries. A good account of the theory and practice of Cox’s model is given by Kalbfleisch & Prentice. Non-parametric Methods are the main focus in the book by Rupert Miller. The Society of Actuaries are understood to have commissioned Norman Johnson to write a study note on Survival Analysis adapted to the needs of actuarial students. This is in an advanced state of preparation.

12.5 Time Series

David Wilkie has written several papers on the use of Time Series models in an investment context.

12.6 GLIM

GLIM has become an industry with its own literature. The GLIM manual and
GLIM newsletters are indispensable for students and practitioners. Two text books associated with GLIM are also important in the literature of statistics. The more elementary is "An Introduction to Statistical Modelling" by Annette J. Dobson. This explains the rationale for fitting maximum likelihood by way of iteratively re-weighted least squares. The more advanced is "Generalized Linear Models" by McCullagh and Nelder. This makes available a wealth of new techniques well illustrated by means of examples. Likelihood concepts are discussed with some notable refinements such as "Quasi-Likelihood Functions".

12.7

Bayesian Methods

The book by Lindley is still the classic exposition. But Bayesian methods are mentioned in many modern books not mainly concerned with a Bayesian approach.

12.8

Decision Theory

A new book by the President of the Institute of Actuaries has been reviewed in a recent JRSS.

12.9

General Insurance — Risk Theory

The book by Beard, Pentikainen & Pesonen is familiar having been in the Institute's reading list for some time. The new third edition is about twice the size of the second edition. Much new material is included and there are many more realistic examples relevant to insurance. One interesting new item of methodology is the use of Time Series ARMA methods in connection with classes of insurance where there is a cyclical variation apparent.

12.10

General Insurance — Loss Distribution
Loss Distributions by Hogg & Klugman achieves a good synthesis between theory and practice. The example of hurricane data has attracted a lot of attention as it has already become a classic example for illustration of the fitting of distributions to the sizes of insurance claims.

12.11

General Insurance - Solvency of Insurers & Equalization Reserves
A two volume work with this title has been produced by a team of Finns. The first volume gives a survey of what has been done to regulate insurers in this respect in several countries. The second volume concentrates on the risk theoretic model.

12.12

General Insurance - Basic Statistical Methods
An introductory textbook has been produced by three Australians: "Introductory Statistics with Applications in General Insurance" by Bossack, Pollard and Zehnwirth. This book (reviewed in JIA Vol. III) is very elementary. It begins by describing logarithms and square roots. But it makes accessible to actuarial students important practical techniques and some advanced thinking on non Experience Rating and Estimation of Outstanding Claim Provisions. There is a tendency for academics remote from industry to disparage "cook books" and to despise textbooks that concentrate on worked examples. But actuarial students have to rely on "self-study" without hints and help from teachers in a classroom and often without useful contact with others studying the same subject. So a book of this kind is valuable when it keeps in view insurance applications and explains how statistical techniques can be used. The use of textbooks on statistics that seem irrelevant to insurance invites working actuarial students to abandon their studies as so many do. Of course a textbook that properly expounds statistical concepts in relation to insurance would be much more valuable.
13. Post Qualification Courses

Once an M.Sc. is set up the prospects for organising Post Qualification courses would be enhanced. Several topics mentioned in the paper are suitable subjects for 1-day, 3-day or 1-week courses. The 1-day course on Box-Jenkins Forecasting required initially the coordination of lecturers from 4 different universities and colleges (and 1 from a computer manufacturer). When an M.Sc. is established in one university a corps of lecturers familiar with the subject matter will be more readily available and the existence of an M.Sc. will also bring about the sort of connections and cooperation with industry that will foster short courses.

14. Acknowledgements

My thanks are due to Tony Puzey for letting me use his notes on the Reverse Yield Gap and to all those who made comments on an earlier version of the paper either at the seminar at The City University on 28th May 1985 or by way of correspondence or in conversations.
REFERENCES


10. Analysis of Survival Data by Cox & Oakes. Chapman Hall.


22. An Introduction to Statistical Modelling by Annette J. Dobson. Chapman Hall.


