A STATISTICAL STUDY OF
THE DISTRIBUTION OF ALCOHOL CONSUMPTION
AND CONSEQUENT INFERENTIAL PROBLEMS

by

John B.F. Field
B.Sc. (Hons.) (Adel.)

Thesis submitted for the Degree of
Doctor of Philosophy
in the University of Adelaide
Department of Statistics

AWARDED 8 MAY, 1986

Adelaide

June 1985

CONTENTS

PART II

SUMMARY

The thesis consists of two parts. Part I examines the distribution of alcohol consumption (that is, the distribution of individual consumers of alcohol according to their consumption averaged over a suitable time period), in relation to Australian data, while Part II considers some more general inferential problems raised in Part I.

After a review of the literature concerning the distribution of alcohol consumption, Part I presents a detailed review of the controversial Ledermann model, providing a new interpretation of some of Ledermann's work. A substantial body of quantitative Australian data is collected together, and then other models, notably various lognormal distributions, are examined in the light of this data. It is found that the most commonly used model of the distribution of alcohol consumption, the two parameter lognormal, spuriously uses information about the light drinkers to make inferences about the heavy drinkers, because of the symmetry of the distribution on the logarithmic scale.

Part II examines this apparent paradox, and suggests some possible solutions. This is done using linear functionals of the class probabilities ("contrasts"). These linear functionals have considerable utility in precisely quantifying important inferential questions, and the mathematics necessary to use them is established. The approach is then to decompose a linear functional to show that a nonparametric estimator of a contrast is partitioned into the parametric estimator plus a second component whose expected value is zero if we can assume the validity of the specification. If we have some

doubt as to the validity of a particular aspect of the parametric specification, we may modify it and so transfer a further component to the parametric estimator, and be confident that the new reduced second component has zero expectation.

We show that, in the case of inferences concerned with the upper tail of the distribution of alcohol consumption, modifying the two parameter log-normal by the addition of a third parameter, or altering the fitting procedure by censoring the lower class frequencies, may ensure valid inferences.

Finally we present a method for fitting a probability distribution subject to a constraint on a linear function of the fitted class probabilities.

SIGNED STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, the thesis contains no material previously published or written by any other person, except where due reference is made in the text of the thesis.

I consent to the thesis being made available for photcopying and loan if applicable if accepted for the award of the degree.

J.B.F. Field

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Professor A.T. James, for his continued assistance, guidance and encouragement throughout the course of many discussions. Dr B.S. Hetzel originally suggested the topic of the thesis, and I have benefitted from discussions with him, particularly in the early stages of the work. In the later stages, discussions with Dr W.N. Venables have been of great assistance. To all three I offer my thanks.

I gratefully acknowledge the support of the CSIRO Division of Mathematics and Statistics, and of its successive Chiefs, Dr J.M. Gani and Dr T.P. Speed.

Finally, my sincere thanks to my wife, Rosemary Field, without whose whole-hearted encouragement and support this thesis might never have been completed.

Chapter 1

Introduction.

> "As a nation Australians drink too much alcohol. The death, sickness, social disruption and economic loss which results has become an unacceptable burden and urgent methods are needed to reduce consumption."

This is the opening paragraph of a *Statement on alcohol consumption and abuse* published by the Australian Medical Association in 1980 (AMA, 1980). The statement, and the policies proposed to reduce consumption, are predicated on the AMA's claim that

> "In any alcohol consuming population, the proportion that are drinking hazardously varies directly with changes in per capita consumption."

In such claims lay the genesis of this thesis. The initial brief was to investigate the background to similar claims, largely based on European and North American data, and see if they were supported by available Australian data.

In the process of this investigation, it became apparent that there were problems of statistical inference underlying some of the basic assumptions. These problems extended beyond the alcohol consumption framework in which they were posed.

The thesis has therefore been divided into two parts. The first part considers the original question of the distribution of alcohol consumption, while the second part raises more general inferential problems and offers some solutions. We give here a brief summary of the two parts.

Part I surveys the literature about the distribution of alcohol consumption, considers various appropriate models and examines them in the light of Australian data. By the distribution of alcohol consumption is meant the distribution of individual consumers of alcohol according to their consumption averaged over a suitable time period. Part I consists of five chapters:

- Chapter 2 surveys the origins of the distribution of consumption in the early 1950s, and traces the changing attitudes to it through the following thirty years.

- Chapter 3 examines the model proposed by Sully Ledermann in 1956, which has been the controversial basis of most subsequent work. A new interpretation of parts of Ledermann's work is given.

- Chapter 4 looks at other models which have been used for the distribution of consumption, notably various lognormal distributions, and the gamma distribution.

- Chapter 5 brings together a large majority of the existing quantitative Australian data on the distribution of alcohol consumption. It is necessary first to consider methods of measuring individual alcohol consumption, and their validity.

- Chapter 6 examines this data, in particular for evidence of the relationship between the "proportion that are drinking hazardously" and "per capita consumption". It is found that the most often used model of alcohol consumption, the two parameter lognormal distribution, spuriously uses information about light drinkers to estimate the number of heavy drinkers. It is shown empirically that by censoring the lower tail of the lognormal distribution, or by adding a third parameter to the distribution, the effects of the lower tail on the estimation of the upper tail can be reduced.

It is this paradox of the light drinkers affecting the estimation of the number of heavy drinkers that is the *raison d'être* for Part II, which considers mathematically the empirical solutions adopted in Part I. Part II consists of one chapter.

- Chapter 7 is concerned with various inferences arising from the fitting of a statistical distribution to grouped data. Initially, the necessary linear algebra is given to enable the later precise formulation of answers to important inferential questions. It is shown that if a distribution is fitted to an observed relative frequency vector, functionals of the fitted probability vector express aspects of the inference assuming the chosen parametric specification. Functionals of the deviations of the fitted probability vector from the observed relative frequencies will express aspects of the goodness-of-fit. By consideration of partitions of these functionals, it is shown that by introducing a third parameter into the two parameter lognormal distribution, or, almost equivalently, by censoring the lower tail of the distribution, the dependence of the upper tail of the distribution on the lower tail can be reduced.

PART I

Chapter 2

The distribution of alcohol consumption – an historical overview.

2.1 Ledermann's original proposals, 1956

The idea of a distribution of alcohol consumption was first put forward by the French demographer Sully Ledermann. He produced a large, two volume report, covering many aspects of alcohol and alcoholism in France (Ledermann, 1956, 1964a). In one of the twenty-three chapters, he dealt with measures of the degree of alcoholic intoxication of a population, considering two such groups: those measures derived from consumption data, and those derived from various alcohol-related diseases. The former group included the number of excessive drinkers in a population,

> "... d'où le problème préalable de la répartition des individus selon leur consommation."

Ledermann proposed that the logarithm of the alcohol consumption was normally distributed, asserting that this was frequently the case for phenomena which develop according to a mechanism of the 'contagion' or 'snowball' type. He quoted several data sets to support his hypothesis.

The basic assumptions of Ledermann's model are

i. alcohol consumption is distributed according to a two parameter lognormal distribution, and

ii. there is a small but constant proportion of drinkers whose daily consumption exceeds one litre (789 g) of absolute alcohol.

Ledermann determined this proportion empirically by pooling estimates from his several data sets.

The assumption of lognormality is unexceptional. However the second assumption generates a relationship between the two parameters of the log-normal distribution, and means that the distribution can be determined by one of them. In particular, there is a relationship between the mean consumption and the prevalence of excessive users. This relationship has considerable implications for alcohol control policies: if the model always holds, an increase in mean consumption will be accompanied by an increase in the number of heavy consumers, and, *vice versa*, it would be possible to reduce the number of heavy consumers by reducing the mean consumption.

Because of the historical importance of Ledermann's contribution to the subject area, the "Ledermann model" is considered in some detail in the next chapter of this thesis. It is shown there that the model is a reparameterisation of a two parameter lognormal distribution. However, in the literature, there has been considerable misinterpretation of Ledermann's work, most of which stems from the failure to distinguish clearly between the model itself, the procedure which Ledermann used to fit the model, and the data which he used to estimate the parameters of the model.

Suppose the usual location and scale parameters of the lognormal distribution, on the logarithmic scale, are μ and σ. Ledermann's reparameterisation of the distribution was in terms of a, the reciprocal of σ, and θ, the standard normal deviate corresponding to a value D on the original scale. Ledermann preset D to a value of 789 g (one litre) of absolute alcohol per day, although he recognised that it was really an extra parameter of the model. It is shown in the next chapter that the choice of a value of D is not critical to the prediction of heavy consumption from the Ledermann model, provided it is large compared with the mean.

The procedure which Ledermann used to fit his model had two stages: taking several data sets he used a graphical technique to estimate a value of θ for each one. He took a weighted mean of these θ values as his estimate of θ. In doing this, he effectively fitted a lognormal distribution to each data set and estimated the proportion greater than D. The standard normal deviate corresponding to the proportion greater than D in the final model is then equal to the weighted mean of the standard normal deviates of the proportions greater than D in the several data sets. The second stage of the procedure required knowledge of the mean (or median) of the population about which predictions were to be made, and it was used to derive a value of the second parameter, a. It is to be noted that if there is only one data set, this procedure implies that the Ledermann model is identically the lognormal distribution of best fit to the data. This has not been recognised in the literature.

Skog (1977a) has summarised the background to Ledermann's assumptions, and is worth quoting in some length:

> "Ledermann's hypothesis of lognormality was in part inspired by the work of the French economist Gibrat (c.f. Ledermann, 1953), who was able to show that a number of economic and social phenomena could be described by the lognormal model. Gibrat (1931) explained this fact through a mechanism which is called the law of proportional effects, later referred to as the snowball effect by Ledermann.
>
> A second source of inspiration was Ledermann's own studies of differences between the French departments with respect to death rates (Ledermann, 1952a). For some diseases, such as cancer, the distribution of the 90 departments with respect to death rates was bell-shaped and in accordance with the Gaussian normal distribution. For contagious diseases, such as T.B., the distributions were highly skewed and approximately lognormal. Arguing that alcohol consumption is a contagion-like phenomenon, Ledermann (1956) consequently concluded that the distribution of consumption should be close to lognormal.

> The second basic assumption in Ledermann's theory is, however, more speculative. It is unclear what made Ledermann believe that the distribution was of such a nature that it left the theoretical percentage with a consumption above 365 litres of pure alcohol per year constant. Ledermann does not substantiate this hypothesis, and there is considerable doubt as to whether it can be given any rational substantiation whatsoever (Skog, 1971).
>
> It seems likely that Ledermann's second assumption is just a way of imposing restrictions on the dispersion parameter, and thereby to generate a relationship between the mean consumption of a population and its prevalence of heavy consumers. That such a relationship should actually exist was not blind guesswork, however. Through his work prior to 1956, Ledermann had come to recognise a close relationship between mean consumption and several indices of harmful effects (Ledermann 1946, 1948, 1952a, 1952b, 1952c, 1953; Ledermann and Tabah, 1951), and this was taken to indicate a relationship between mean consumption and prevalence of heavy users."

The second assumption has not only been criticised, but has been constantly misinterpreted in the literature. It finds expression in various forms; of those listed below, only the first is correct.

> "the theoretical proportion above 365 litres annually ... can be considered constant and identical in all populations." (Skog, 1982)
>
> "a lognormal distribution with a fixed limit" (Cartwright, Shaw and Spratley, 1978b)
>
> "one percent of the population consume in excess of one litre of absolute alcohol per day" (Duffy, 1977b)

But this is not the only source of confusion. It has been widely assumed in the literature (Skog 1971, 1973, 1977a, 1980a, 1982, 1983; Smith 1976a, 1976b; Duffy 1977a, 1977b, 1980; Duffy and Cohen 1978; de Lint 1974; Cartwright, Shaw and Spratley, 1977, 1978b; Miller and Agnew 1974; Singh 1979; Tuck 1980; Furst 1983) that Ledermann intended the value of θ which he had determined should be kept constant and used in all future applications. However a careful reading of Ledermann (1956) shows that this was not the case. Ledermann regarded θ as a parameter of his model, and the value he determined for θ (= 3.43) as an estimate of the true value "if it exists". This view will be amplified in the next chapter. Thus in any application of the model, the estimate of θ (and of a) should be determined from the available data.

Certainly the data Ledermann used to estimate the parameters of his model was inadequate in several respects, and it has given rise to many objections (Skog, 1971, 1982; Miller and Agnew, 1974; Smith, 1976a; Parker and Harman, 1978; Tuck, 1980). Some of his samples were small and unrepresentative, and some were clinical; the details of data collection procedures were sketchy; the samples contained both consumption and blood alcohol concentration (BAC) figures. (Ledermann also showed that two sets of alcohol sales data could be fitted by a three parameter distribution, but he did not use them in his determination of θ.) Undoubtedly at that time it was a matter of using what data he could find.

In 1964 Ledermann published a second volume which continued his earlier work, and in particular, lent support to his hypothesis of a relationship between the mean and heavy consumption. (Ledermann, 1964a). Using data from additional sources (Brezard 1958, 1959, 1960) he plotted the proportion of consumers of 10 cl or more per day against the average consumption, repeating the plot for consumers of 20 cl or more per day. On these graphs he also plotted the theoretical curves generated by his model. Although Skog (1973) has pointed out that for populations with small differences in mean consumption there are some anomalies in the sample points, over a wide range of mean consumptions there was close agreement between the empirical points and the theoretical curve. However Ledermann did not use the Brezard data to test his lognormal hypothesis; Skog (1980b) has demonstrated that four of the seven data sets show significant deviation from lognormality.

It is interesting that Ledermann's original volume (1956) was ignored in the literature for some twelve years. Schmidt and Popham (1978) have suggested that this neglect was because alcohol researchers in the late 1960s denied any role to the overall level of consumption in a population as a determinant of the prevalence of heavy use. They distinguish two schools of thought at that time relevant to alcoholism prevention, the first deriving from the classical disease concept of alcoholism, where an alcoholic was believed to differ fundamentally from social drinkers. In this "bimodal model" (Popham, Schmidt and de Lint, 1976) the distribution of consumption would be bimodal, and factors influencing the consumption of normal drinkers will have little or no effect on the consumption of problem drinkers. This lead to treatment of alcoholics as the main remedial measure. The second school noted that in some European countries where alcohol is used regularly with meals and is an integral part of everyday activities, gross drunkenness and other types of dangerous drinking appeared to be uncommon. They advocated that alcohol be 'demystified' and made more generally available, with prevention being achieved by encouraging drinking as an incidental part of routine activities. Popham, Schmidt and de Lint (1976) have called this model the "integration model". Neither school could see any use for a distribution of overall consumption. Ledermann's work, and that of others to be discussed below, provided a third model of prevention, the so-called "single distribution" or "unimodal" model.

2.2 The period 1968 – 1975

In 1968 then, Jan de Lint and Wolfgang Schmidt of the Addiction Research Foundation (ARF), Toronto, Canada, showed that alcohol consumption in Ontario, as measured from retail sales of wine and spirits, closely followed the Ledermann model. Despite criticisms of the work, it was largely this paper which brought Ledermann's work to the attention of alcohol research workers outside France (Edwards, 1973). The criticism has concerned the distribution of purchases not necessarily being the same as the distribution of consumption (Skog, 1971, 1973), and the absence of any statistical testing of goodness-of-fit (Miller and Agnew, 1974). More recently, Duffy (1977a, 1977b) thought that the Ledermann distribution had been incorrectly fitted; this was refuted by Skog (1980a).

Between 1968 and 1975 the main thrust of research into the Ledermann model was concentrated in Toronto and in Scandinavia; this latter work was principally in Oslo, at the National Institute for Alcohol Research, but also at several institutions in Helsinki. This period culminated in the publication of a report (Bruun *et al*, 1975) as the result of a collaborative project of the Finnish Foundation for Alcohol Studies, the WHO, and the ARF in Toronto. The report purported to be a "state of the art" paper, and presented, *inter alia*, a concensus view on the distribution of alcohol consumption. We now consider the research which led up to this report.

The Toronto group were enthusiastic about the Ledermann model. They produced tables to facilitate calculations (Hyland and Scott, 1969; de Lint, 1974). Smart and Schmidt (1970) fitted lognormal distributions to BACs obtained in several earlier studies (Holcomb, 1938; Lucas *et al*, 1953;

Vamosi, 1960; McCarroll and Haddon, 1962; Borkenstein *et al*, 1964) of vehicle drivers not involved in accidents, claiming that the distribution fitted the data well. This was not surprising, since three of the five data sets contained only three class intervals. O'Neill and Wells (1971) subsequently showed that the only data set with a reasonable number of class intervals (Borkenstein *et al*, 1964) showed significant deviation from the lognormal distribution. More importantly they pointed out that reducing the mean BAC would not necessarily reduce the proportion of impaired drivers, as had been stated by Smart and Schmidt, since changes in the dispersion parameter had been ignored. Ekholm (1972) gave approximate formulae for evaluating the change in the proportion with changes in the mean and standard deviation of the population.

Schmidt and de Lint (1970) compared four methods of measuring the prevalence of alcoholism in Ontario: consumption data using the Ledermann model, deaths from alcoholism, from liver cirrhosis and from suicide, finding reasonable agreement in all cases. They concluded that estimation of alcoholism prevalence from consumption data was the most practical, since all the necessary data was relatively easy to obtain.

At a Symposium on Law and Drinking Behaviour in 1970, Schmidt stated that Ledermann's distribution had been shown to apply to data in various countries, with differing attitudes to drinking, beverage preferences, drinking habits, legislative controls and educational efforts.

> "Our conclusion is that, for all practical purposes, the *form* of the distribution is unalterable and of such a character that *excessive consumption is inextricably linked to general consumption*" (Popham, Schmidt and de Lint, 1971; emphasis in original).

At the same Symposium, Room criticised the Ledermann model, mainly on the grounds of using cross-sectional data to establish "the manner in which changes must inevitably occur." (Room, 1971).

De Lint and Schmidt (1971b) encapsulated the philosophy of the single distribution model at that time:

> "Since rates of alcoholism rise and fall with the overall level of alcohol use in a population, a reduction in *per capita* alcohol consumption must lead to lower rates of alcoholism."

They used the Ledermann model to calculate estimated rates of those drinking in excess of a daily average of 15 cl of absolute alcohol for 21 countries. Similar figures, taken from de Lint (1974) were used by the WHO Expert Committee on Drug Dependence (WHO, 1974).

In another series of publications, Smart and co-workers demonstrated that a lognormal distribution was, in most cases, an adequate description of summary scores for frequency of use of a wide range of drugs. (Smart, Whitehead and Laforest, 1971; Smart and Whitehead, 1972, 1973; Smart, 1978; Castro, Chao and Smart, 1978). These studies were largely on students; McDermott and Scheurich (1977) found similar results in a telephone survey of residents of Kansas. In contrast to most alcohol consumption data, the drug use scores are based on frequency of use, and little attention appears to have been given to the effect that the construction of the score may have on the distribution. The 1973 paper of Smart and Whitehead is notable in that it marks a moving-away from the strict assumption of a lognormal distribution:

> "It may well be that that the unimodal, continuous character of the distribution is more important for prevention than the presence or absence of lognormality."

Schmidt (1973) outlined some of the difficulties encountered in using sales and consumption data to estimate the magnitude of alcoholism, including illicit production, tourism and changes in stocks. Problems in international comparisons of alcoholism included the variation in coverage reported by governments, and differing *per capita* consumption by alcoholics in different countries. He also presented evidence that under-reporting of consumption in surveys was not equal at all levels of consumption, being much greater at high levels. Schmidt also noted that use of *per capita* consumption as an index of rate of excessive use of alcohol depended on the assumption that alcoholism could be defined as the consumption of a fixed quantity of alcohol, and gave evidence to support this.

At this time too one of the earliest applications of the Ledermann model to Australian data was made (Rankin, 1971). James Rankin of the ARF used the model to present trends in the number of heavy drinkers, assessing that their incidence had increased by 56 percent between 1949 and 1968. He gave a greater exposition of the implications of Ledermann's work while delivering the Seventh Leonard Ball Oration in January 1974.

One of the first of the Scandinavian workers in the field was Klaus Mäkelä, of the Social Research Institute of Alcohol Studies, Helsinki. He reported that annual alcohol consumption in interview surveys of a representative sample of the adult Finnish population in 1968 and 1969 approximated a lognormal distribution in both years, although he found small deviations in the upper tail (Mäkelä, 1969, 1971a, 1971b, 1971c). He also noted that because the lognormal distribution is continuous and unimodal, there is no objective way to define a population of alcoholics in terms of consumption (Mäkelä, 1971a).

Ole-Jørgen Skog in Oslo was among the first to critically examine the Ledermann model (Skog, 1971). He pointed out that Ledermann's own data did not support his hypothesis that the proportion of the population consuming in excess of 365 litres per year was constant and independent of the mean consumption. Skog also recognised that the Ledermann model would overestimate the number of heavy consumers, particularly in countries of low mean consumption. In a subsequent paper (Skog, 1973b) he concluded

> "I would ... like to emphasise that I do not insist that his [Ledermann's] conclusion is faulty. My point is, rather, that there is no foundation in the available material for a conclusion that categorical. This is why I feel Ledermann's assertion to be more of a hypothesis with some foundation, rather than a well-documented conclusion."

Skog's approach was rather to fit two parameter lognormal distributions to consumption data, and then look for a relationship between the two parameters to effectively eliminate one parameter, as he perceived Ledermann's aim (Skog, 1971, 1974). This approach was subsequently taken up by Bruun *et al* (1975).

Skog fitted two parameter lognormal distributions to eight data sets (seven reported alcohol consumption and one for BACs) and found significant deviations from lognormality in four of them. He noted that these deviations, and those reported by Mäkelä (1969, 1971a) indicated that a less skewed distribution than the lognormal might give a better fit; he fitted the gamma distribution to his data, concluding that it fitted all except one of the data sets (Skog, 1974).

Skog also tried using a three parameter lognormal distribution, as Ledermann had suggested for heterogeneous data, but with little success - he achieved an acceptable fit in only one out of five data sets. But his later conclusions (Skog, 1977a) do not wholly reflect this: he says the three parameter lognormal distribution, as an answer to the heterogeneity problem

> "... is a very unfortunate solution however. Firstly the third parameter has no significant theoretical interpretation. Secondly a three-parametric distribution is so flexible that it can be fitted to almost all empirical data. Thirdly the relation between mean consumption and prevalence of heavy consumers is destroyed (within the theory, that is)."

Another critical assessment of the Ledermann model was made by Miller and Agnew (1974). They considered the usefulness of the model from the points of view of description and prediction, and had severe reservations in both instances. Their objections included the problems of determining the mean consumption; verification that a population was homogeneous; the equivocal nature of empirical validation; the lack of verification on longitudinal data; the tacit assumption that all alcoholic beverages were equally implicated in alcoholism. They concluded

> "The evidence suggests that the distribution of consumption is probably distributed according to a positively skewed distribution such as the lognormal distribution. It is not likely, however, that consumption is distributed exactly as hypothesized by Ledermann. ... At the predictive level the usefulness of the model is even more in doubt."

Several later critiques of Ledermann's original proposals can be mentioned. Smith (1976a, 1976b) examined the model from the point of view of prediction of heavy consumers, and was highly critical, describing the procedure as "an example of bad statistical methodology." Parker and Harman's (1978) consideration of Ledermann's work is unique in that they state that Ledermann's parameter θ is probably a variable rather than a fixed quantity. Only in the special case of a population known to be homogeneous in drinking practices and culture will the lognormal model be dependent on only one parameter. In reply to Parker and Harman's paper, Schmidt and Popham

(1978) show the extent to which the ARF group had moved away from their earlier position espoused in such papers as de Lint and Schmidt (1971a, 1971b) and de Lint (1974). They (Schmidt and Popham, 1978) state

> ".. constancy ... in the relationship between mean and dispersion is not a prerequisite. ... it is not essential whether the distribution belongs to the lognormal family, the gamma family, or some other class of distribution."

This reflects the thinking contained in the monograph of Bruun *et al* (1975), which we will now consider.

2.3 The report of Bruun *et al*, 1975

This report, entitled *Alcohol Control Policies in Public Health Perspective*, was prepared by a working group from Finland, Norway, UK, USA and Canada. It describes alcohol-related health damage, trends in alcohol consumption and production, and the need for policies which place high priorities on alcohol availability.

As a "state of the art" paper it is notable for the cautious stand it takes in comparison with the papers before it. Smith (1976b) has suggested that this may be a consequence of the co-operative nature of the report, but given the evidence of papers such as Smart and Whitehead (1973), this seems unlikely. The authors propose a considerable dilution of the Ledermann model. The basic approach owed much to Skog (who was a co-author), largely following from his 1971 and 1974 papers. The distribution of consumption is said to be a highly skewed distribution, described by "two main parameters, the mean and a measure of dispersion." The lognormal distribution is mentioned only in two examples.

Prompted by criticism of the work, Skog (1980a) has summarised the approach of Bruun *et al*:

> "What Bruun *et al* tries to demonstrate, is the existence of a relationship between mean consumption and prevalence of heavy use - not in the *strong* sense, but in the *weak* sense."

Skog's use of strong and weak sense relationships refers to differing interpretation of the hypothesis of covariation between *per capita* consumption and prevalence of heavy use. The first interpretation ("strong") is that

> "populations with identical mean consumption levels have close-to-identical prevalence rates, too"

and the second interpretation ("weak")

> "populations with highly different mean consumption levels are likely to have different prevalence rates, too"

The effect of these differing interpretations is that the weak relationship

> ".. would enable us to offer statistical predictions of the effect of large changes in mean consumption levels with respect to prevalence rates, but it does not imply the possibility of obtaining an estimate for the prevalence rate in a given population on the basis of *per capita* consumption alone."

The latter possibility is seen as a consequence of a strong relationship.

In support of their weak sense relationship, Bruun *et al* use two empirical justifications. By plotting the standard deviation of the logarithm of consumption against the mean consumption for data from six adult and eight youth samples, they show that "differences as to dispersion between populations with similar levels of consumption are quite small." They conclude that the "apparent stability in dispersion seems to indicate a certain invariance in the distributional pattern." "Invariance" is left undefined, but it appears to be used in a non-statistical sense. The rationale behind the figure is that if a substantial increase in total consumption should fail to lead to an increase in the prevalence of heavy consumers, then we would observe a considerable decrease in the dispersion parameter. Thus a change in the mean consumption will generally be an expression of a collective movement of the entire population upwards or downwards along the consumption scale (Skog, 1983).

This justification has been criticised by Smith (1976a, 1976b), de Lint (1976), Duffy (1977a, 1977b, 1980) and Duffy and Cohen (1978). Skog (1983) has agreed with much of this criticism, particularly the limited data base on which the diagram is based, the fact that the plotted points represent highly significant differences in dispersion, and the implausibility of precisely representing a distribution by its first two moments. Despite these objections, Skog (1983) maintains that difference in total consumption actually is an expression of difference along the entire consumption scale. Using data from twenty-four samples he regresses several percentiles of consumption against average consumption, both on a logarithmic scale. The regression line for each percentile (25th, 50th, 75th, 90th and 95th) shows a positive slope, on which fact he bases his conclusions. In another objection, Duffy and Cohen calculated the dispersion and *per capita* consumption from a survey of Scottish drinking habits (Dight, 1976) and found that the values for female drinkers did not follow the pattern suggested by Bruun *et al.* Skog (1980a) has suggested that this is the result of a methodological artifact, but does not present a convincing argument.

The second empirical justification of Bruun *et al* is a diagram reproduced from Skog (1971), relating proportion of heavy consumers (more than 10 cl/day) to *per capita* consumption. The data used in this diagram includes that used by Ledermann (1956, 1964a, 1964b). The plotted points are derived directly from data without recourse to fitted distributions, and show an approximately quadratic relationship. The authors conclude that over the range of mean consumption which is of practical importance, substantial differences in heavy consumption are evident, and that substantial changes in mean consumption are likely to be accompanied by substantial changes in the number of heavy consumers.

Having established this relationship, the authors defend it against the charge that most of the data are cross-sectional rather than longitudinal: "... in this instance it is hard to see why a longitudinal study should produce results significantly different from those found by cross-cultural comparisons." They point out that residual variation in the fitted regression line is small in spite of the wide divergences among the populations under review as to drinking pattern and sociocultural characteristics, so that these factors must have little effect on the proportion of heavy consumers. They also quote two longitudinal studies to support their stand (Eckholm, 1972, who uses data from Mäkelä, 1971b; Brun-Gulbrandsen, 1976). In both these studies, the mean consumption increases with time.

Their final conclusions are as follows:

> "1. A substantial increase in mean consumption is very likely to be accompanied by an increased prevalence of heavy users.
>
> 2. If a government aims at reducing the number of heavy consumers this goal is likely to be attained if the government succeeds in lowering the total consumption of alcohol."

Most of the literature before the publication of this monograph was concerned with using the fitted distributions, be they called Ledermann or lognormal, for prediction of the numbers of heavy drinkers. As has been stated, this was Ledermann's original aim. But in deserting the "strong" sense relationship between mean consumption and heavy use for a "weak" one, Bruun *et al* have in fact altered the purpose of their inference: they are now more concerned with the "possibility of making *statistical* predictions as to the effect of large changes in *per capita* consumption" (Skog, 1980a; emphasis in original). Much of the current literature has still not caught up with this change of purpose.

2.4 The period since 1975

Changes to the single distribution theory since 1975 have been lesser in extent and degree than changes in the previous decade. In contrast to the shift of emphasis which was the most notable modification of that period, there has been

- a gradual acceptance of the position of the "weak" relationship between mean consumption and heavy use, although this acceptance has been marked by polemic discussion in the literature;
- an attempt at an explanation of the weak relationship in terms of social interactions, in contrast to the previous reliance on empirical justification;
- discussion of control policies based on the single distribution theory, and finally
- a continuing stream of papers reporting surveys of consumption, examining survey methodology issues and so on.

While all four of these items are inextricably linked, it is the first one which is our principal present concern.

In January 1977, a symposium on "The Ledermann Curve" was held in London at the invitation of the Alcohol Education Centre. Six papers were presented giving current thoughts from members of the Canadian, Scandinavian and British groups. This symposium makes a convenient organising point for this section of the review, and we shall trace various paths leading from it.

In an initial overview paper, de Lint (1977) described the consumption curve as continuous, unimodal, positively skewed and probably lognormal, but was hesitant about using it to estimate the prevalence of excessive use.

> "And, in any event, it would seem more useful to investigate the current increases in consumption, their effects on public health and how these trends can be stabilised than to produce estimates of excessive use."

In his paper, Skog (1977a) not surprisingly takes the line of Bruun *et al* (1975) as far as the distribution of consumption is concerned. That is, the distribution is "approximately lognormal" with an apparent "invariance" in the distributional pattern. He suggests tackling the problem of aggregation of subpopulations by replacing the one general distribution with a system of distributions, one for each level of aggregation of the population. This system, he suggests, could be based on the gamma distribution, with parameters related by empirically determined constants.

As has been mentioned earlier, in his 1974 paper Skog had shown that the gamma distribution often gave a better fit to consumption data than did the lognormal distribution. At the London symposium, he hypothesised that the lognormal distribution could be a correct choice of model if consumption was determined by a large number of multiplicative factors all contributing a small, equal amount to the total variance. Aitchison and Brown (1954, 1957) have shown that the aggregation of several lognormal subpopulations will be lognormal if the variance of each subpopulation is constant, and if the number of subpopulations is large enough for the distribution of mean consumptions of each subpopulation to be both continuous and lognormal. However, Skog considered that some factors contributed a *large* amount to the variance tending to make the distribution less skewed than lognormal. He therefore proposed the gamma distribution, which was in accord with studies showing empirical distributions to be somewhat less skewed than lognormal (Mäkelä, 1969; Skog, 1971, 1974).

Guttorp and Song (1977) reanalysed data from Skog (1971) showing that the gamma distribution gave a reasonable fit in only one out of six cases, whereas the lognormal distribution fitted the data well in five cases. This has been disputed by Skog in an exchange of views (Skog, 1979a; Guttorp and Song, 1979; Skog, 1979b) concerning methods of fitting distributions and testing goodness-of-fit, culminating in a claim by Skog that the test of separate families of hypotheses for discrimination between lognormal and gamma distributions (Cox, 1961, 1962; Jackson, 1968, 1969) gives biased results with grouped data. It is notable that the dispute concerned the goodness-of-fit of distributions to the whole of each data set, not to the fits in the tails.

Skog (1980b) fitted both lognormal and gamma distributions to data from Brezard (1958, 1959, 1960). This was the "confirmatory" data used by Ledermann (1964a). The data, a random sample of the general population from seven districts in France, are exceptional in that under-reporting of consumption appears to be very small (Skog, 1980b; Brezard, 1958). Skog's results were inconclusive. He has suggested that this may be largely caused by the aggregation of data from both sexes in Brezard's data. He showed (1977b), by comparison of a number of male and female populations having similar consumption levels, that systematic differences appeared to exist between the distributional pattern of the sexes. Admitting his conclusions to be highly tentative because of a modest data base, he found a "somewhat smaller" prevalence of heavy use among females than among males.

In his London symposium paper, Skog also outlines a theory of social interaction between persons to try to explain some of the empirically established facts of the distribution. He later expanded this theory (Skog, 1979c, 1980c). Another model to explain increasing consumption patterns was set out at the same symposium by Sulkunen (1977). This model was based on aspects of the drinking practices of the population.

A notable paper at the "Ledermann Curve" symposium was given by John Duffy, a statistician with the M.R.C. Unit for Epidemiological Studies in Psychiatry, Edinburgh. Duffy (1977a) was highly critical of the single distribution approach, and it is a pity that the impact of some of his criticisms have been lessened by other criticisms based on a series of misunderstandings and misinterpretations. The paper firstly misinterprets the prevailing thinking about the Ledermann model, stating that the model assumed "one percent of the populaton consume in excess of one litre per day". The misinterpretation persists in Duffy (1977b) and Duffy and Cohen (1978). Duffy's symposium paper also misinterprets the work of Bruun *et al* (1975). All this led to a series of papers (de Lint, 1978; Skog, 1980a; Duffy, 1980) which at times descended to petty point scoring, but which have had positive aspects as well: considerable clarification of the approach of Bruun *et al* (1975) came from Skog's (1980a) paper (by way of the explicit distinction between the strong and weak hypotheses), and there have emerged some genuine criticisms of research in the distribution of alcohol consumption.

Duffy's main sustainable criticisms have been

i. that distribution theorists have been using the one fitted distribution for several purposes: description, estimation, and testing of empirical theory.

ii. that when goodness-of-fit testing has been attempted, no account has been paid to the fact that the degree of fit in the centre of the distribution may have little bearing on the degree of fit in the tails.

iii. comparisons between populations are of little value in considering the effect of changes within populations.

But other of Duffy's criticisms amount to statistical hair-splitting. To maintain that "the distribution cannot be continuous, because the populations are not infinite", as he does in his 1978 paper, is to ignore the fact that it is real data with which we are dealing, not some theoretical example. In the light of a lack of evidence as to any discontinuities in the distribution, continuity is a reasonable assumption to make for mathematical and conceptual convenience, and one which, it need hardly be said, is often assumed in sociological and biological situations.

In a recent paper, Duffy (1982) maintains that there is

> "no such thing as the distribution of alcohol consumption: there are as many distributions of alcohol consumption as there are populations of consumers."

He does, however, concede that for some purposes (estimation of means and variances, and hypothesis testing) the logarithm of alcohol consumption may be assumed to be normally distributed. His position seems well summarised by the following quotation.

> "The empirical distribution of alcohol consumption between respondents in a survey is an essential part of modelling and estimating relationships involving alcohol consumption. However, the investigation of goodness-of-fit of a particular mathematical form for the empirical distribution should proceed from consideration of its effect on the conclusions of the analysis. Studies which consider fitting mathematical distributions *in vacuo* are of little value, and this is

> especially true when it is the tails of the distribution which are of particular interest."

An interesting and somewhat different approach to the whole problem of alcohol consumption levels has been taken by Taylor (1979). he has suggested recasting the distribution curve of consumption as a table of rates, akin to the use of mortality rates. This rate he terms a "consumption containment rate" (CCR), as it measures, at any level of alcohol consumption, the number of drinkers (per thousand, say) not consuming a further unit of alcohol. Taylor points out that the CCR at any level of consumption is independent of the CCR at low and medium levels, although the sampling variance will be considerable in the tail end of the distribution. The CCR facilitates easy comparisons between different studies. Taylor demonstrates that the lognormal distribution is characterised by a decreasing CCR at high levels of consumption, implying that the tendency to refrain from having another drink falls as consumption increases. He suggests that the property of a relatively steady or declining CCR as consumption increases may be a better characteristic of alcohol consumption data than a lognormal distribution. Additionally it may be possible to calculate CCRs for individuals and relate them to CCRs derived from population data.

2.5 Discussion

Ledermann originally proposed that the distribution of alcohol consumption was lognormal, with a relationship between the two parameters of the distribution implying a relationship between mean consumption and prevalence of excessive users. Ledermann's work will be considered in more detail in the next chapter, and so we defer detailed discussion of it until then. But despite widespread criticism of his work, and misinterpretation of the details of it, there is still a strong body of support for his final conclusion: to achieve a reduction in alcoholism and alcohol-related problems, it would appear necessary to substantially reduce the mean *per capita* consumption (Ledermann, 1956, p 159; 1964a, p 430).

Ledermann's model postulated a near-quadratic relationship between the mean consumption and the proportion of excessive consumers. When new data confirmed this relationship, he considered it "un fait de la plus grand importance" (Ledermann, 1964a, p 443). This empirical relationship is the common ground between Ledermann's original model and the more recent point of view, such as has been espoused by Bruun *et al* (1975), de Lint (1977), Smart (1977), Schmidt and Popham (1978), Cartwright, Shaw and Spratley (1978b), and Skog (1980a). This concensus has alcohol consumption distributed in a continuous, unimodal, positively skewed manner, "similar to" a lognormal distribution.

We can note in passing that this means there are no clear distinctions between categories of drinkers - light drinkers merge into moderate drinkers, and that class in turn merges into heavy drinkers. Any boundaries between classes are artificial, and are there only to categorise a continuous situation.

It is hard to disagree with the concensus view of the distribution, or to find contradictory evidence. Recent reports of populations as diverse as rural Punjab males (Mohan *et al*, 1980), residents of Vancouver (Storm and Cutler, 1981), New Zealand adolescents (Stacey and Elvy, 1981) and North Sea oil rig workers (Aiken and McCance, 1982) all support the view.

However it is the implications of the distribution for alcohol control policies that are more controversial. To put the present discussion into perspective, we list some of the available control policy models. A convenient typology is provided by Robinson (1982), who distinguishes preventative strategies on the basis of what is perceived as the central focus of "the alcohol problem": alcoholics, society, or alcohol itself.

1. *Focus on alcoholics*: based on the bimodal model of consumption, the major effort is put into treatment and support of those individuals identified as incurring social costs - the "alcoholics".

2. *Focus on society*: based on the integration model, the mass media and education systems are used to disseminate information giving guidelines for healthy drinking, and encouraging responsible use of alcohol.

3. *Focus on alcohol*: based on the single distribution model of consumption, the aim is to reduce *per capita* consumption through regulating price and availability of alcohol.

All policy models have factors which operate against them (Mandell, 1982). In terms of the classification above, these include

1. cultural resistance to labelling individuals as alcoholics; legal difficulties in applying sanctions; cost of operating treatment centres.

2. the cost; competing values in society; competing information in the media and education system.

3. economic benefits of increased consumption to farmers, producers, distributors and governments. Bruun *et al* (1975) recognised that the reduction of such benefits may be perceived as outweighing any benefits to be gained from the application of the control policies.

Among those who advocate control measures based on the single distribution theory, there are differences of opinion about the effect of the policies, which are related to whether the distribution model is interpreted in the "strong" or "weak" sense of Skog (1980a). For instance, the (British) Advisory Committee on Alcoholism (1977, quoted by Tuck, 1980) states that "measures which raise or reduce the overall level of drinking result in a corresponding increase or decrease in the number of harmed individuals". And in Australia, the AMA (1980) stated

> "In any alcohol consuming population the proportion that are drinking hazardously varies directly with changes in per capita consumption."

These are similar to Ledermann's conclusion of twenty years earlier, mentioned above. Skog (1981) takes a more conservative line:

> "*large* changes in per capita consumption are likely to result from similar changes in consumption among drinkers at all levels, and the prevalence of heavy use is therefore *likely* to go up."

And the single distribution view is certainly not without critics. For example, Tuck, in a controversial paper (1980), considers it

> "neither necessarily correct nor helpful, indeed it may stand in the way of more promising and flexible policies."

It is not within the scope of the present study to advocate any particular policy. However, the propriety of using the empirical relationship between mean consumption and excessive use as a basis of such policies can at best be regarded as questionable. The relationship, such as is given in Ledermann (1964a, 1964b) and Bruun *et al* (1975) is derived from cross-sectional studies, rather than longitudinal ones, and is not directly applicable to the prediction of change within the one population. It is reasonable however to hypothesise on the basis of the relationship that a reduction in average consumption will lead to a reduction in excessive use, but experimental evidence of such a reduction is needed before control policies could be soundly advocated.

Mäkelä (1978) points out that it is not easy to find recent examples of a decreasing level of alcohol consumption. He resorts to the indirect indicators of health and criminal statistics to show that in several situations early this century, where it is known that *per capita* consumption declined, there was a corresponding decline in the indicator statistics. He points out that other concurrent factors, such as war or popular mass sentiments favouring temperance, make it problematic to generalise from such historical experiences. He also considers evidence from various liquor strikes, where reduced availability of liquor is the only influence on consumption. In a conclusion qualified because the database is from countries where drinking has not been integrated with everyday social life, he says

> "The evidence ... seems to indicate that the decrease has been accompanied by diminished intake among heavy drinkers and by a reduced frequency of obnoxious drinking occasions."

And even in a climate of increasing consumption, there appear to be few recorded longitudinal studies measuring the effects of increased consumption on the one population. Bruun *et al* (1975), as was mentioned earlier, quoted two Scandinavian studies in support of their argument. More recently Cartwright (1977) and Cartwright, Shaw and Spratley (1977, 1978a, 1978b) examined data from two surveys in 1965 and 1974 of a South London suburb, and concluded that a change in the total consumption of the population was associated with a change in the prevalence of alcohol-related problems.

Implicit in policies dependent on the single distribution theory is the belief that the prevalence of heavy users is closely correlated with various public health problems. There is a considerable literature on the subject which we have not attempted to review here. However in a review publication, Moser (1980) agrees that the conclusions of Bruun *et al* (1975) on this subject, namely

> "heavy users of alcohol have a substantially elevated risk of premature death"

and that

> "the aetiological importance of alcohol is clear with respect to deaths from cirrhosis of the liver",

are widely accepted as being based on a fairly reliable mass of data. The possibility exists however that some alcohol-related problems are related to a particular drinking pattern rather than to heavy use *per se*.

Finally we may ask, given the concensus view on the distribution of consumption, what is the relevance to control policies of fitting mathematical distributions to consumption, sales or BAC data? We have already stated the need for studies in situations of decreasing alcohol consumption. Such studies need to involve careful analysis of relative frequencies in the extreme upper tail, and for this purpose a parametric fit based primarily on relative frequencies in the middle and low upper tail has advantages over the use of raw data involving small numbers of observations. However the choice of a parametric specification needs careful consideration, and should depend on what aspects of inference are involved. A specification which is satisfactory for one purpose may be quite unsuited for another. We will return to this problem later of this thesis.

Chapter 3

The Ledermann model of alcohol consumption.

3.1 Introduction

The French demographer Sully Ledermann (28 Oct. 1915 - 1 Mar. 1967) appears to have been the first to have put forward the idea of a distribution of alcohol consumption, in 1956. He was concerned, *inter alia*, with estimating the number of excessive drinkers in the French population,

> "... d'où le problème préalable de la répartition des individus selon leur consommation."

Ledermann's basic proposition was that the logarithm of the alcohol consumption was normally distributed. While this assumption has been acceptable, Ledermann's method of fitting his model has been the subject of much discussion.

Because there has been so much discussion and, we believe, misinterpretation of Ledermann's work, the model is considered here in some detail. This does not imply, however, that we advocate its use.

3.2 The Ledermann procedure and the Ledermann model

3.2.1 Overview To aid the understanding of Ledermann's work, it is helpful to distinguish between the "model" which Ledermann used, and the "procedure" or process by which the model is constructed.

The *Ledermann procedure* is a method for combining several samples, or "subpopulations", each assumed to be distributed according to a two parameter lognormal distribution, to obtain a "pooled" lognormal distribution for the entire population. The subpopulations are combined in such a manner that the standard normal deviate corresponding to the proportion of the pooled population greater than some preset consumption level, D, where D is large compared with the mean consumption, is equal to the weighted mean of the standard normal deviates corresponding to the subpopulation proportions greater than D. This pooled distribution, the *Ledermann model*, is a reparameterisation of a two parameter lognormal distribution. If we have only one subpopulation then the Ledermann model is identically the lognormal distribution of best fit.

3.2.2 Description Suppose we have some "target" population for which we wish to estimate a distribution of individual alcohol consumption. We assume we have some estimate of the mean* of this population. We suppose that we have available to us several samples of alcohol consumption data, each sample coming from some subpopulation of the target population. The *Ledermann procedure* enables the combination of information in the subpopulations to give the *Ledermann model* for the target population.

---

* Ledermann also gave an equivalent method for deriving the model from the median, rather than the mean, of the target population.

We assume we have k samples, and let X_i be the variable representing the consumption of the ith subpopulation. We suppose that a two parameter lognormal distribution with parameters μ_i and σ_i can be fitted to each sample:

$$X_i \sim \text{LN}(\mu_i, \sigma_i) \qquad\qquad i = 1 , \ldots , k.$$

Choose a value of D, large relative to the mean. Ledermann called D the "maximal consumption", and defined it as "la consommation approximative très rapidement mortelle". D is really a parameter of the model, but Ledermann took it to be preset at a fixed value. For each sample we calculate the standard normal deviate θ_i

$$\theta_i = \frac{\log D - \mu_i}{\sigma_i} \qquad\qquad i = 1, \ldots , k \tag{3.01}$$

Thus if U is a standard normal variate,

$$\Pr\,(X_i > D) = \Pr\,(U > \theta_i)$$

is the predicted proportion of consumers in the ith subpopulation with consumption greater than D, and θ_i is the standard normal deviate corresponding to this proportion.

We then calculate a weighted mean of the θ_i values to give the first parameter, θ, of the Ledermann model

$$\theta = \frac{\sum_{i=1}^{k} n_i \theta_i}{\sum_{i=1}^{k} n_i} \tag{3.02}$$

where n_i is the sample size for the ith subpopulation. This step is the basis of the combination of the subpopulations, and with the assumption that the target population is also distributed as a two parameter lognormal, determines a family of lognormal distributions whose parameters μ and σ are related by

$$\theta = \frac{\log D - \mu}{\sigma} \qquad (3{:}03)$$

All members of this family have the property that the standard normal deviate corresponding to the proportion greater than D is equal to the weighted mean of the standard normal deviates of the proportions greater than D in the subpopulations.

To choose one of this family as the Ledermann model, we use our knowledge of the mean consumption, m, of the target population. Since m is an estimate of ξ, the mean of the lognormal population, we can write

$$\log m = \mu + \tfrac{1}{2}\sigma^2 \qquad (3.04)$$

We can solve (3.03) and (3.04) for σ. Eliminating μ from the two equations leads to a quadratic equation for σ, which will have real roots if

$$\theta^2 \geqslant -2 \log \left(\frac{m}{D}\right) \qquad (3.05)$$

Ledermann defined his second parameter to be*

$$a = \frac{1}{\sigma} \qquad (3.06)$$

and expressed the quadratic equation in terms of a rather than σ. He took the larger root, giving

$$a = \frac{\theta + \sqrt{\theta^2 + 2 \log \left(\frac{m}{D}\right)}}{-2 \log \left(\frac{m}{D}\right)} \qquad (3.07)$$

In the event that real roots did not exist, Ledermann took

$$a = \frac{1}{\sqrt{-2 \log\left(\frac{m}{D}\right)}} \qquad (3.08)$$

(An explanation of the second root, and the situation of complex roots, will be given in Section 3.4).

---

* Ledermann (1956) gives his derivation using logarithms to both base 10 and base e, using the lower case a and the upper case A respectively for the same parameter. We use the lower case a in line with the usual statistical notation. Strictly, it corresponds to Ledermann's A.

The Ledermann model is then specified by the two parameters θ and a, and we may write

$$X \sim \text{LED}\,(\theta, a \mid D)$$

3.2.3 Summary To fit the Ledermann model, the steps in the Ledermann procedure are

i. choose a value of D.

ii. fit a two parameter lognormal distribution to each sample, and calculate the θ_i values using (3.01)

iii. calculate θ using (3.02)

iv. calculate a using (3.07) or (3.08) as appropriate

and the Ledermann model is

$$\text{LED}(\theta\,,\, a \mid D)$$

3.3 The Ledermann model as a reparameterisation of the two parameter log-normal distribution

From the above description, it is easily seen that the Ledermann model is just a reparameterisation of a two parameter lognormal distribution. For, by (3.06),

$$\sigma = \frac{1}{a} \tag{3.09}$$

and then by (3.03)

$$\mu = \log D - \frac{\theta}{a} \tag{3.10}$$

and we have the equivalence

$$\mathrm{LED}(\theta\ , a \mid D) \equiv \mathrm{LN}\,(\log D - \frac{\theta}{a}, \frac{1}{a}) \tag{3.11}$$

That is, the Ledermann model with parameters θ and a is a two parameter lognormal model with parameters $(\log D - \frac{\theta}{a})$ and $\frac{1}{a}$.

The main misconception in the literature about the Ledermann model is that the value of θ, once determined by Ledermann, was to be taken as a fixed constant for all times. From the formulation above (equation 3.11), it is obvious that if θ is regarded as fixed, the distribution depends only on the one parameter, a. We will return to this point later.

If we have a sample from only one subpopulation (i.e. $k = 1$), and this sample is fitted by a two parameter lognormal distribution, that is

$$X_1 \sim \mathrm{LN}\,(\mu_1, \sigma_1)$$

then by equation (3.02) we have $\theta = \theta_1$. Substituting this value in equation (3.07) and using $m = \exp(\mu_1 + ½\sigma_1^2)$ we find, after a little algebra, that $a = \frac{1}{\sigma_1}$. Substituting for θ_1 from (3.01) in (3.10) produces $\mu = \mu_1$, i.e. the Ledermann model is $\mathrm{LN}\,(\mu_1, \sigma_1)$ and is identically the lognormal distribution of best fit to the original sample.

3.4 Characterisation

In any given application of the Ledermann procedure, once θ has been determined from the θ_i values, a family of lognormal distributions is determined. The parameters are related by equation (3.03), which we can write as

$$\mu = \log D - \sigma\theta \ .$$

This family can be represented by a line on a graph of σ against either μ or ξ, the mean consumption. Figure 3.1 shows, as a function of σ and ξ, the line generated by D = 789 g/day and θ = 3.43, i.e. the values used by Ledermann. The points **M** and **F** represent the particular distributions he chose for his predictions for male and female heavy consumers in France. In terms of Figure 3.1, these distributions were selected by reading off from the graph values of σ corresponding to ξ and then calculating μ from the relation $\xi = \exp(\mu + ½\sigma^2)$.

Smith (1976a) has represented the family of lognormal distributions by a pencil of straight lines through the point (θ , D) on logarithmic probability paper. This is equally valid, but we prefer the conciseness of the present representation and the ease of presenting comparisons of different families.

Figure 3.2 shows the different families of distributions produced by varying the value of D, taking θ to have the value 3.5. Figure 3.3 shows the different families produced by various values of θ, for a fixed value of D = 700 g alcohol/day. Figures 3.4 and 3.5 reproduce the same information, but in terms of μ and σ rather than ξ and σ. From Figure 3.2 it might appear that the value chosen for D will have a large effect on the final Ledermann model. However it should be remembered that if the value of D is altered, then by equation (3.01) so will the subpopulation values of θ_i,

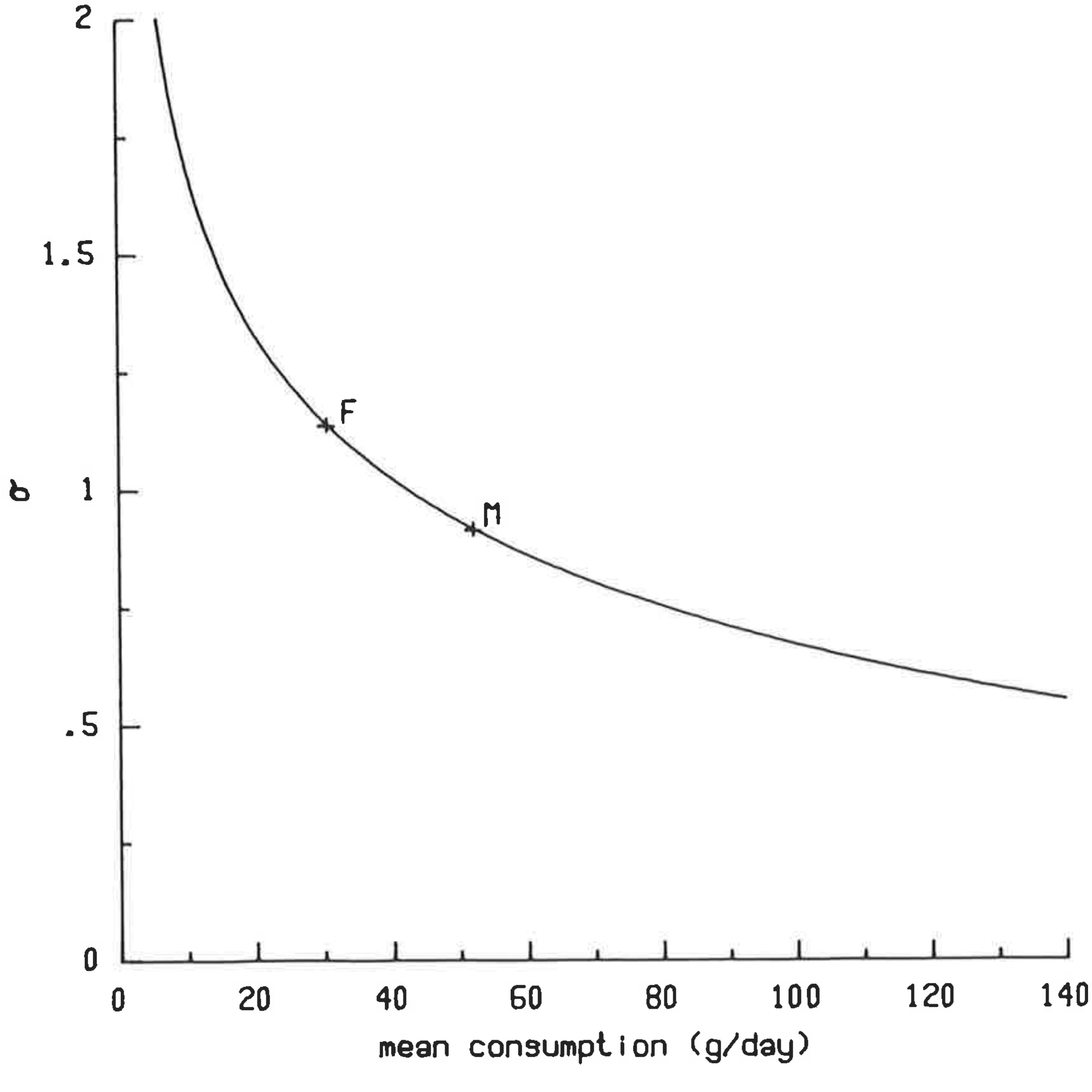

Figure 3.1 The family of lognormal distributions used by Ledermann (D = 789 g/day, θ = 3.43), and his models for males (M) and females (F).

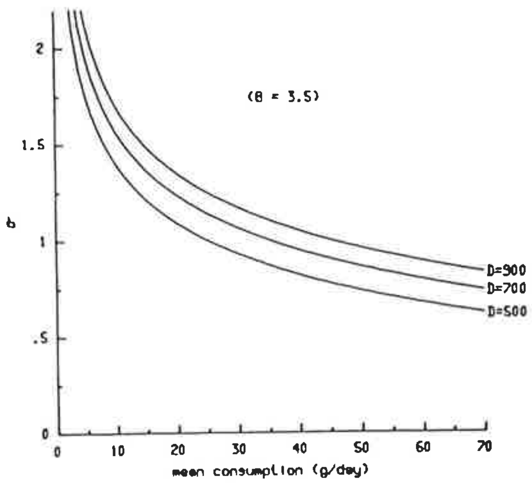

Figure 3.2

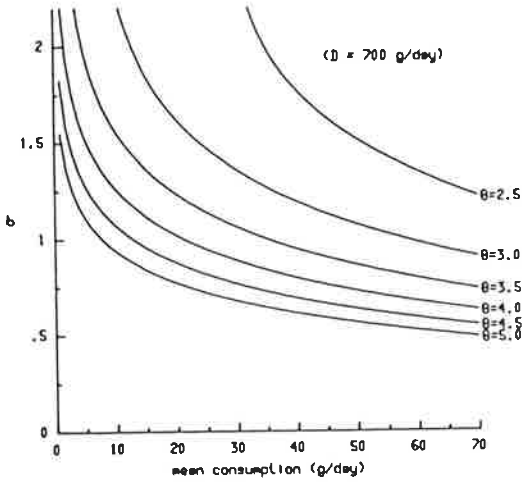

Figure 3.3

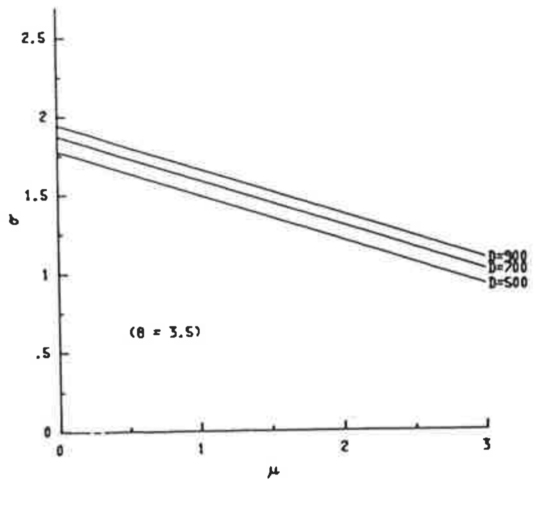

Figure 3.4

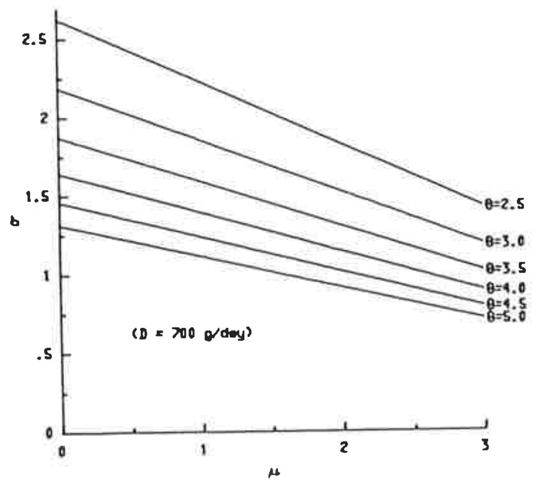

Figure 3.5

Figures 3.2 – 3.5 Families of lognormal distributions generated by the Ledermann model with differing values of D and θ, shown as functions of σ and either the mean consumption or μ.

which will in turn change the value of θ. Hence when it comes to choosing a particular lognormal distribution as the Ledermann model, the family with which we are dealing will have altered in a manner indicated by Figure 3.3. We will examine the effect of the choice of D further in Section 3.7.

Figures 3.2 and 3.3 show the family of distributions over the usual ranges of the mean consumption and σ. If we consider the shape of one of the curves as σ increases beyond this range, we find that rather than asymptoting to the y axis, the curve is "U" shaped, the tangent to the base of the "U" being parallel to the y axis. Thus corresponding to any mean consumption, there are two values of σ as solutions. The second value, on the upper arm of the "U", is the smaller root of the quadratic equation corresponding to equation (3.07). The case of complex roots occurs when the mean of the target population lies to the left of the vertical tangent to the base of the "U". Ledermann's solution to this case (equation 3.08) is to take σ equal to the value at which the tangent touches the base of the curve.

3.5 Ledermann's data

Ledermann's reason for considering the distribution of consumption was to enable prediction of the numbers of heavy drinkers in France. Therefore it would seem appropriate that he take as his subpopulations for determining his pooled value of θ, a reasonably balanced cross-section of the total French population. This was not the case however. Undoubtedly his choice was dictated by what was available but it has been widely criticised (for example Miller and Agnew, 1974; Smith, 1976a, 1976b; Parker and Harman, 1978; Skog, 1982).

It is instructive to superimpose on Figure 3.1 the points representing the subpopulations which Ledermann used.* This is done in Figure 3.6, where circles represent alcohol consumption data, squares represent blood alcohol content (BAC) data, and the figures by each point are the weights (= sample size) expressed as percentages, used to combine the values of θ_i. The equivalence between the consumption and the BAC scales is given by

$$\frac{\text{consumption}}{D_{\text{consumption}}} = \frac{\text{BAC}}{D_{\text{BAC}}}$$

that is, in this case

$$\frac{\text{consumption}}{789} = \frac{\text{BAC}}{0.4} .$$

The weights indicate the extent to which BAC data dominates the fit, supplying 83% of the information for the determination of θ. In particular the sample from the Chicago car drivers (Holcomb, 1938) has a weight of 63%. Thus the distribution Ledermann derives could more appropriately be called a dis-

---

* The points plotted on Figure 3.6 are from lognormal distributions fitted to each of Ledermann's data sets. Ledermann did not estimate the parameters of each subpopulation, but used a graphical technique to estimate the θ_i values directly. Using fitted distributions to calculate the θ_i gives a value of θ = 3.31 instead of the 3.43 found by Ledermann.

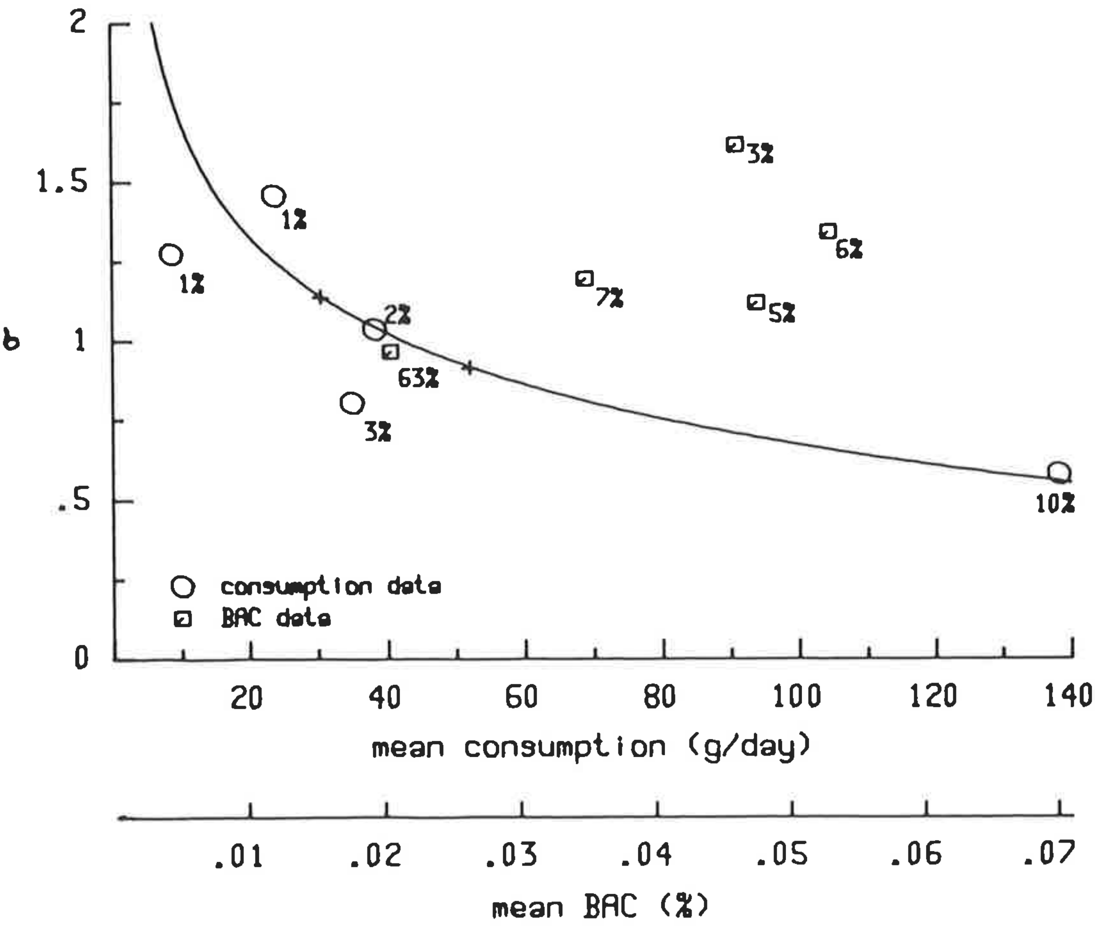

Figure 3.6 The data Ledermann used in his calculations. The percentages are the weights given to each sample in determining the weighted estimate of θ.

tribution of BAC levels than a distribution of alcohol consumption. The wide scatter of points indicates the diversity of the populations he used.

The quality of Ledermann's data should, however, be seen to reflect only on the reliability of his estimates of the number of heavy drinkers in France in 1954, not on the methodology used to obtain the estimates.

3.6 The value of θ

In introducing D into his argument, Ledermann assumed (p 262)*

> "... que l'intervalle de consommation allant de 0 à D contienne une proportion F_D des consummateurs, cette proportion étant supérieure à 99% par exemple."

In a footnote, he explains that this is a common statistical convention when fitting a distribution of infinite range to data of finite range. This statement has been misunderstood by several researchers. For example, Duffy (1977a, 1977b) and Duffy and Cohen (1978) interpreted it to mean "one percent of the population consume in excess of one litre per day" until corrected by Skog (1980a), while Cartwright, Shaw and Spratley (1977, 1978b) thought it meant that the endpoints of the distribution were fixed.

But the more general misinterpretation of the method concerns the status of Ledermann's value of θ. He determines the individual θ_i values for each of his data sets and uses their weighted mean as his "provisional" estimate of θ for "general" calculations (p 275):

> "Nous adopterons provisoirement, pour les calculs généraux, la valeur θ = 3.43, c'est-à-dire F_D = 99.97%".

This statement has been misinterpreted to mean that the proportion F_D = 99.97% (and hence $1-F_D$ = 0.03%) and the value θ = 3.43 should be taken as fixed for all applications. For example, see de Lint and Schmidt (1968); Skog (1971, 1980a, 1982); de Lint (1974); Smith (1976b); Singh (1979).

But Ledermann did not intend θ to be fixed at this value. In fact he regards θ as one would regard any parameter in a model: it has a "true" value, and in fitting the model to a particular data set, one determines an estimate of this true value from the data. He even admits the possibility

---

* In this chapter, page references are references to Ledermann (1956).

that a true value might not exist. Having assembled the various estimates of θ from his data sets into Table 5.1.6, he says (p 275)

> "Ce sont là des estimations, dont l'écart par rapport à la valeur "vrai" – si elle existe – dépend de plusiers facteurs: nombre d'observations disponibles, conditions de formulation des réponses, etc."

Ledermann then calculates the weighted mean of the θ values, and takes that as his "provisional" estimate of the true value of θ.

A little further on, he is quite explicit about this. Referring to his Table 5.1.7 which gives predictions for proportions of heavy consumers in populations with various mean consumptions, he says (p 275)

> "Soulignons que les indications données par cette table sont théoriques, relatives à une population homogène, et découlent des valeurs D = 365 litres d'alcool pur par an et θ = 3.43 adoptées.
>
> Pour une distribution concrète, il faudra ajuster la répartition normale-logarithmique correspondante, selon la méthode classique: D peut rester le même, mais θ peut alors varier, comme l'ont montré les exemples dont nous avons déduit une estimation moyenne θ = 3.43."

which my translation gives as

> We emphasise that the information given by this table is theoretical, relating to a homogeneous population, and follows from adopting the values D = 365 litres of absolute alcohol per year and θ = 3.43.
>
> For a particular distribution, it will be necessary to fit the corresponding lognormal distribution by the standard method: D will remain the same, but θ will vary, as has been shown in the examples from which we have deduced an estimated mean value of θ = 3.43.

As further evidence, consider the predicted standard normal deviate corresponding to a consumption level, X. In our notation

$$U = \frac{\log X - \mu}{\sigma}$$

Substituting for μ and σ from (3.09) and (3.10) gives

$$U = a \log\left(\frac{X}{D}\right) + \theta \qquad (3.12)$$

Ledermann gives this equation (equation (42), p 275) and remarks

"Pour une distribution concrète, a et θ sont déterminés directement."

In contrast, de Lint (1974, Appendix II) calls equation (3.12) the "Ledermann equation", "in which θ = 3.43". To add further confusion to the subject, Parker and Harman (1978) quoted de Lint's formulation of this equation, but wrongly labelled U as "Student's t distribution", and $a \log(\frac{X}{D})$ as "the log transformation of consumption for the distribution", both meaningless terms in this context. However they are at least correct in their assessment of θ: they say

> "For Ledermann θ is not a given but a variable whose value, at best, is approximated through the use of weighted means".

However the mainstream of the literature has continued to regard θ as fixed at the value of Ledermann's estimate.

3.7 The value of the maximal consumption

Ledermann based his procedure on three "known" data points

i. the zero consumption, 0

ii. the mean consumption, m

iii. the "maximal consumption", D.

In introducing D into his argument he says (p 262)

> "Nous la prendrons égale à 100 cl. d'alcool pur par jour, soit 365 litres d'alcool pur par an. Un quatrième paramètre conventionnel, qui enleve son importance physique au chiffre retenu de 100 cl doit être associé à cette limite."

(Ledermann initially began with a three parameter lognormal distribution with parameters μ, s and w, where w is the usual threshold parameter. He disposed of w by setting it equal to zero.) However he did not use D as a parameter to be obtained from the data, but rather fixed it in advance of his calculations. For consumption data, Ledermann took D equal to 100 cl/day (789 g/day), and for BAC data he took D equal to 0.4% (4 g °/oo), although for one data set (p 273) he gave values of θ calculated using D = 0.5% as well.

Ledermann gives no justification for his choice of levels of D other than, for consumption (p 262),

> "D ... est la consommation approximative très rapidement mortelle."

and for BACs (p 271)

> "Nous prendrons pour limite D, une alcoolémie de 4 g °/oo, alcoolémie à partir de laquelle commence les accidents mortels."

and (p 103)

> "Les accidents mortels commencent à une alcoolémie de 4 g. p. 1000 environ..."

In the literature there has been little discussion of Ledermann's choice of D for BAC data, although the choice of D for consumption data has aroused much discussion.

Ledermann's reason for fitting distributions was to enable prediction of the number of heavy consumers in France. To judge the effect that choice of the value of D has on this prediction, the data which Ledermann used to determine his value of θ has been used to fit Ledermann models with varying values of D. The predicted proportions of consumers drinking more than 80 g alcohol per day are shown in Figure 3.7, for varying values of D and ξ, the mean consumption. Clearly the proportions are very insensitive to D.

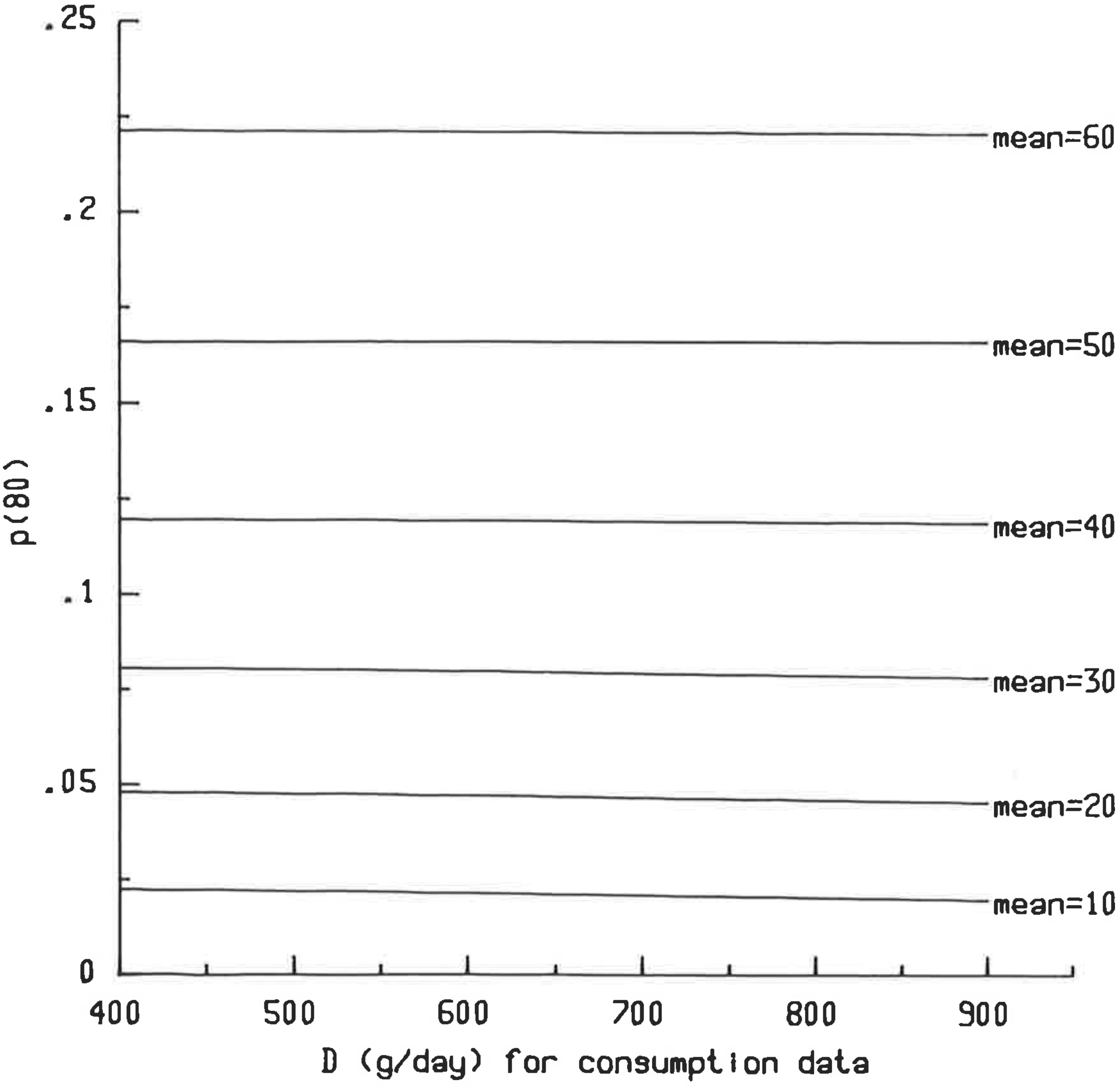

Figure 3.7 Predicted proportions of consumers drinking more than 80 g alcohol/day, for varying values of *D* and ξ, the mean consumption.

3.8 An example

To illustrate the Ledermann procedure, we use data from the 1978 Busselton, W.A. survey (Cullen *et al*, 1980). The data is given in Table 5.24, and includes a breakdown by age and sex. We use the data for males only, with the six age groups forming our subpopulations.

The first step is to fit two parameter lognormal distributions to each of the age groups. Table 3.1 gives details of the fits.

Table 3.1

Two parameter lognormal distributions fitted to the 1978 Busselton male data, and calculations of θ

i	age group	n_i	μ_i	σ_i	χ^2_3	estimated proportion > 80 g/day	θ_i
1	< 30	249	3.0715	.7942	3.04	.0495	4.549
2	30-39	237	2.9315	.7971	2.25	.0344	4.709
3	40-49	205	2.8964	.9437	8.35*	.0577	4.014
4	50-59	209	3.1179	.8837	14.59**	.0763	4.036
5	60-69	202	2.9152	.9822	8.36*	.0677	3.838
6	70+	158	2.7897	.8353	0.48	.0283	4.663

$\sum n_i = 1260$ * ≡ P < 0.05   ** ≡ P < 0.01

We take a value of D = 800 g/day, although we shall see later that this choice is not critical. For each age group we calculate

$$\theta_i = \frac{\log 800 - \mu_i}{\sigma_i}$$

Values of θ_i are shown in Table 3.1. We can then calculate a pooled estimate of θ,

$$\theta = \frac{\sum n_i\theta_i}{\sum n_i}$$

which gives

$$\theta = 4.307$$

The sample mean for the overall population is $m = 27.476$. Using this, we calculate a from equation (3.07)

$$a = \frac{\theta + \sqrt{\theta^2 + 2 \log (\frac{m}{D})}}{-2 \log (\frac{m}{D})}$$

$$= 1.1485$$

from whence we have

$$\sigma = \frac{1}{a} = 0.8707$$

We then calculate μ:

$$\mu = \log D - \theta\sigma$$

$$= 2.9342$$

Thus the Ledermann model is the two parameter lognormal distribution

LN(2.934, 0.871).

The value of θ determines the family of lognormal distributions represented by the line in Figure 3.8, on which are also plotted the points representing the six age groups, and the final model. By comparison of Figures 3.6 and 3.8, it is obvious that the agegroups used in this example form a much more homogeneous group of "subpopulations" than did Ledermann's data, even though three of the agegroups show significant discrepancy from lognormal distributions.

Since Ledermann's purpose was to estimate the number of heavy drinkers, we can examine various estimates of the proportion of drinkers consum-

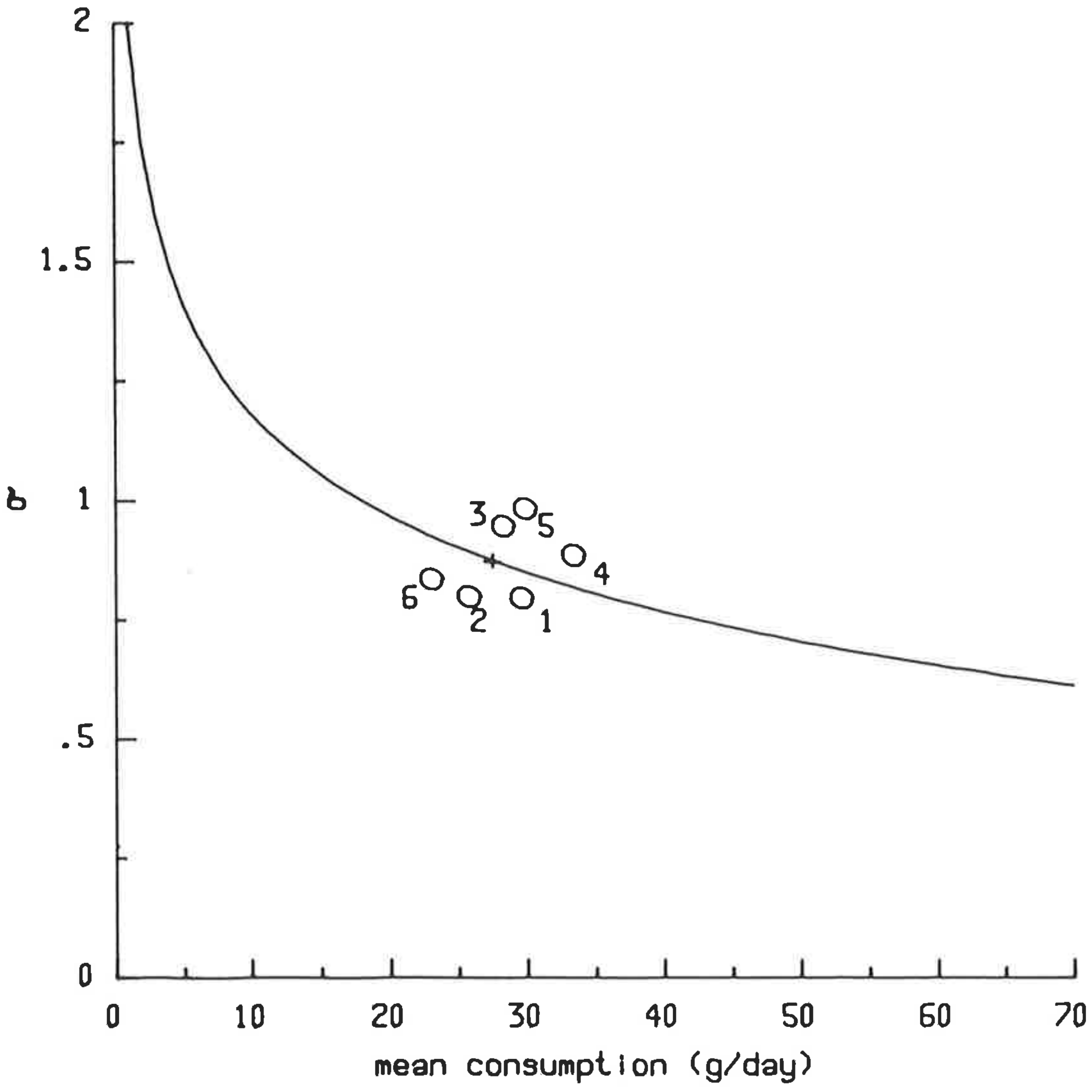

Figure 3.8 Age groups (1 - 6), the Ledermann family of lognormal distributions, and the Ledermann model (+) for the 1978 Busselton males.

ing in excess of 80 g/day. The weighted average of the individual class proportions, given in Table 3.1, is 0.0433, while the Ledermann model gives 0.0482. A two parameter lognormal distribution fitted to the data summed over age groups (LN(2.1649, 0.8759); χ^2_3 = 8.56, P < 0.05) gives 0.0057, while a three parameter fit (LN(3.4862, 0.6353, −13.6224); χ^2_2 = 2.78, NS) gives 0.0487. A two parameter fit censored below 40 g/day (that is, the class intervals below 40 g/day amalgamated into one class) gives 0.0482 (LN(3.0830, 0.7814); χ^2_2 = 2.54, NS). These results are summarised in Table 3.2.

Table 3.2

Estimated percentage of male heavy drinkers, Busselton, 1978

weighted average of age groups	4.33%
Ledermann model	4.82%
two parameter lognormal	0.57%
three parameter lognormal	4.87%
censored two parameter lognormal	4.82%

We shall see in Part II of this thesis that we might expect the "best" estimate to be that given by either the three parameter, or the censored two parameter lognormal fits; indeed these gave non-significant fits to the data. The Ledermann model is in good agreement with these, despite the fact that only half of the subpopulations had nonsignificant lognormal fits. The two parameter distribution gives a gross underestimate, but this is not surprising since the distribution does not fit the data well.

We can examine the effect that our choice of *D* = 800 g/day has had on our estimates. Table 3.3 gives details of the Ledermann models fitted with varying values of *D*. We see that the model is very insensitive to the

Table 3.3

The effect of *D* on estimates of the proportion of heavy drinkers

D (g/day)	θ	σ	μ	proportion > 80 g/day
200	2.702	2.930	0.877	0.0488
400	3.171	2.931	0.874	0.0485
600	3.974	2.934	0.872	0.0483
800	4.307	2.934	0.871	0.0482
1000	4.566	2.935	0.870	0.0481

choice of *D*, which confirms the earlier evidence of Figure 3.7.

However the model is much more sensitive to the value of the mean, *m*. Table 3.4 shows the Ledermann models and predicted changes in the estimated percentage of heavy consumers with changes in the mean, *m*, ranging from 15% below to 15% above its calculated value 27.476.

Table 3.4

The effect of the mean on estimates of the proportion of heavy drinkers

mean		μ	σ	proportion > 80 g/day	
value	% change			value	% change
23.355	−15	2.729	0.918	0.0359	−25.5
24.728	−10	2.802	0.902	0.0398	−17.4
26.102	−5	2.870	0.886	0.0439	−8.9
27.476	0	2.934	0.871	0.0482	0
28.850	+5	2.995	0.857	0.0527	+9.3
30.224	+10	3.053	0.843	0.0575	+19.3
31.597	+15	3.108	0.830	0.0625	+29.7

Quite small changes of 5% of the mean alter the estimated proportion of the population drinking more than 80 g/day by almost twice this percentage. Over the range of change considered here, the ratio remains roughly constant, particularly for positive changes.

3.9 Discussion

It is not difficult to understand why there has been so much misinterpretation of Ledermann's work. It was written in French, and was published by the National Institute of Demographic Studies in France, as a volume of a monograph series which, ten to fifteen years after its publication when researchers outside France became interested in the work, was unlikely to be readily available. Therefore some of the early interpretations of the work were reused by other workers. For example, Singh (1979) quotes de Lint's (1974) description of the "Ledermann equation". Additionally, Ledermann's description is not easy to follow. It extends over fifteen pages, and does not clearly distinguish between the subpopulations and the target population; he started his description with the target population, and then later, when he needed to estimate θ, introduced the subpopulations. Having dealt with this aspect, he then returns to the target population. We believe that drawing a clear distinction between the subpopulations and the target population, as has been done in this chapter, leads to a clearer understanding of the model.

The main misunderstanding has been that the value of θ determined by Ledermann should be taken as a fixed value. We have shown that Ledermann intended his estimate of θ to be just that: an *estimate* of some "true" value θ. Seen in this light, Ledermann's data loses much of its controversial nature, as it reflects only on the quality of his estimates, not on the procedure itself.

The literature contains much discussion of Ledermann's choice of a value of D. His aim in fitting the model was to estimate the number of heavy consumers in France in 1954. Used for this purpose we have seen that the choice of D, provided it is much larger than the mean, is almost irrelevant.

What then can be said about the validity of the model? In using the Ledermann procedure with more than one subpopulation we must make the following assumptions.

a. The distribution of consumption in each of the subpopulations is lognormal

b. The distribution of consumption in the target subpopulation is also lognormal

c. That the proportion greater than D is constant for all populations. In other words, there does exist a true value of θ.

If there is only one subpopulation, then, as has been shown, the Ledermann model is the best fitting lognormal distribution, and we need only assume that the underlying distribution is lognormal.

In his original monograph, Ledermann (1956) made no tests of significance of the fits of his subpopulation lognormal distributions. In fact, as we have already mentioned, he did not explicitly fit distributions to his data, but for each subpopulation, plotted the data on log-probability paper and drew in a line of best fit, probably by visual inspection (Skog, 1982). He then read off the value of θ for the subpopulation directly.

In the example above we have also ignored the fact that for three of the age subpopulations, a lognormal distribution does not give a non-significant fit to the data, and have proceeded to use all the data to fit the Ledermann model. Under these circumstances it is surprising that the predictions based on the Ledermann model and those based on the censored two parameter and three parameter lognormal distributions fitted to the consumption data for all agegroups combined, agree so closely. It would be unreasonable to expect such agreement in all cases where subpopulations showed marked deviations from lognormality. In general, an experimenter fitting the Ledermann model would have to decide if any differences from lognormal distributions among the subpopulations were due to chance fluctuations or to model misspecification, before proceeding to fit the final model.

The most unusual aspect of the Ledermann model is the method of combining subpopulations, via the weighted estimate of θ. As we have shown, this step determines a family of lognormal distributions whose members all have the property that the standard normal deviate corresponding to the proportion greater than D is equal to the weighted mean of the standard normal deviates corresponding to the proportions greater than D in the subpopulations. By itself, the step is not sufficient to determine the Ledermann model uniquely, the extra information needed to do this being supplied by way of the mean (or median) of the target population.

The user of the procedure is required to accept that there exists a "true proportion greater than D", or equivalently, a "true value of θ". For instance, Ledermann's estimate of this true proportion was 0.03%, using D = 789 g/day. By way of comparison, recalculation of θ from the example in Section 3.8 using the same value of D gives the estimate of the true proportion as 0.00089%. Can we accept that these two are the estimates of the same "true" quantity? If we were to increase D towards its limit, the proportions will both approach zero, but from a practical point of view it seems unreasonable to accept that there does exist such a "true" value.

Apart from this concern with the theory behind the model, a practical question to be addressed is whether the Ledermann procedure gives a good method of estimating a distribution of alcohol consumption to be used for prediction of heavy consumption. As to this qualification about use of the model, we maintain that any model should be judged in relation to the use to which it will be put, and that different models may be required to estimate different features from the one data set. Since the estimation of heavy consumption was Ledermann's main use for his model, we feel justified in using this criterion in this case. In this we disagree with Schmidt and Popham (1978) who state

> "It is regrettable that Parker and Harman - and others before them - have been preoccupied with the shortcomings of the Ledermann equation as a device to obtain specific estimates of prevalence."

We have seen in the example that the proportion of heavy consumers derived from the model is very sensitive to changes in the estimated mean of the target population. Since this estimated mean will usually not be known with any great certainty, any estimates of heavy consumption derived from the model will necessarily be suspect. The Ledermann model, with its heavy reliance on this mean, may therefore not be a good model to use for this purpose.

In summary, the model cannot be recommended as a means of estimation of the excessive use of alcohol. However this does not preclude the existence of an empirical relationship between mean consumption and prevalence of heavy use. It is difficult not to agree with Smith (1976a), who in giving evidence to the (Australian) Senate Standing Committee on Social Welfare, says

"The extraordinary thing is I have a good deal of sneaking regard

> for Ledermann's ingenuity and I suspect that he has done us a service in bringing out one or two partial truths."

It is indeed a pity that Ledermann died in 1967 before his work in this area became so widely known. Had he lived, he may have been able to prevent much of the confusion surrounding his work.

Chapter 4

Other models of alcohol consumption.

4.1 The two parameter lognormal distribution

4.1.1 Definition Let X = mean individual alcohol consumption, $0 < X < \infty$. The range of X excludes zero, i.e. we consider only consumers of alcohol, and ignore abstainers.

Let

$$Y = \log X$$

be the logarithm to the base e of X. (We use natural logarithms throughout.) Then if Y is normally distributed with mean μ and standard deviation σ, we say that X is lognormally distributed with mean μ and standard deviation σ, and write

$$Y \sim N(\mu, \sigma)$$

and

$$X \sim LN(\mu, \sigma).$$

The probability density function for Y is then

$$f(y) = \frac{1}{\sigma\surd(2\pi)} \exp\left\{ -\frac{1}{2\sigma^2}(y-\mu)^2 \right\} dy \qquad -\infty<y<\infty$$

and for X:

$$f(x) = \frac{1}{\sigma x\surd(2\pi)} \exp\left\{ -\frac{1}{2\sigma^2}(\log x - \mu)^2 \right\} dx \qquad 0<x<\infty$$

4.1.2 Characteristics The rth moment about the origin is given by

$$\mu_r' = \frac{1}{\sigma\surd(2\pi)} \int_0^\infty \frac{x^r}{x} \exp\left\{ -\frac{1}{2\sigma^2}(\log x - \mu)^2 \right\} dx$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp(ry) \exp\left\{ -\frac{1}{2\sigma^2}(y-\mu)^2 \right\} dy$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma^2}(y^2-2(\mu+r\sigma^2)y+\mu^2) \right\} dy$$

Completing the square in the exponent, and putting $t = y - r\sigma^2$ gives

$$\mu_r' = \exp(r\mu + \tfrac{1}{2}r^2\sigma^2) \cdot \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma^2}(t-\mu)^2 \right\} dt$$

$$= \exp(r\mu + \tfrac{1}{2}r^2\sigma^2)$$

Thus we have the <u>mean</u>, ξ

$$\text{Mean } (X) = \text{E}(X) = \mu_1' = \exp(\mu + \tfrac{1}{2}\sigma^2)$$

and the variance

$$\text{Var } (X) = \mu_2' - (\mu_1')^2$$

$$= \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

$$= \xi^2 (\exp(\sigma^2) - 1)$$

Since μ is the median of Y, the <u>median</u> of X is

$$\text{Median } (X) = \exp(\mu).$$

The distribution is unimodal, with <u>mode</u>

$$\text{Mode } (X) = \exp(\mu - \sigma^2)$$

Thus mean(X) > median(X) > mode(X).

To show how density curves change with changing values of μ and σ, Figure 4.1 shows three density curves for lognormal distributions with $\sigma = 1$ and $\mu = 1, 2, 3$. Figure 4.2 shows three density curves for $\mu = 2$ and $\sigma =$ 0.5, 1.0 and 1.5.

4.1.3 <u>The proportion of heavy consumers</u> Let $p(\ell)$ = proportion of consumers above some limit, ℓ. Then

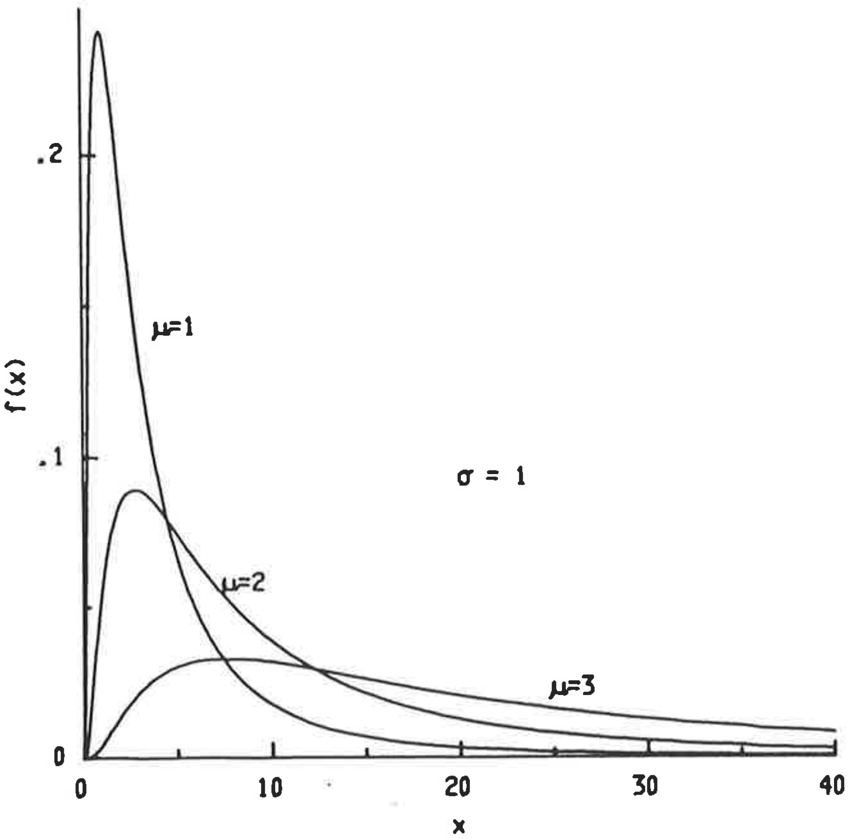

Figure 4.1 Frequency curves of the two parameter lognormal distribution with σ = 1 and three values of μ.

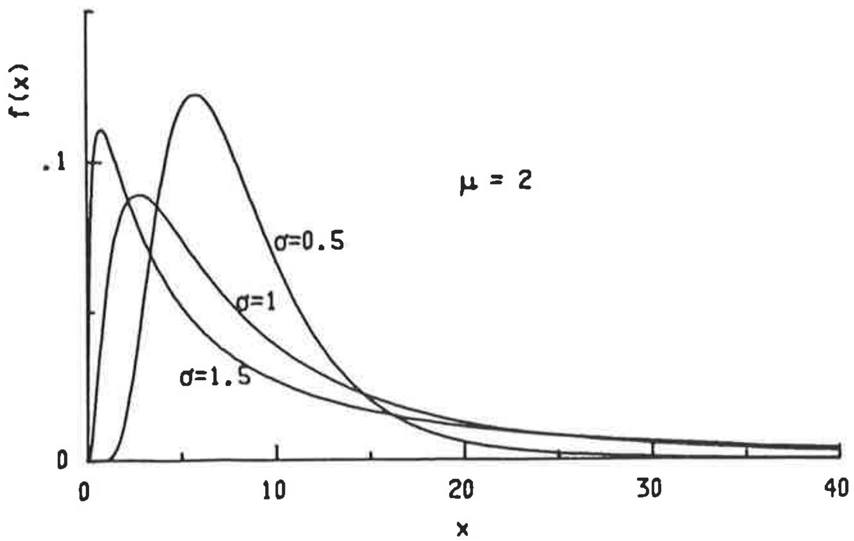

Figure 4.2 Frequency curves of the two parameter lognormal distribution with μ = 2 and three values of σ.

$$p(\ell) = \frac{1}{\sigma\surd(2\pi)} \int_{\log \ell}^{\infty} \exp\left\{ -\frac{1}{2\sigma^2}(y-\mu)^2 \right\} dy$$

$$= 1 - \Phi\left(\frac{\log \ell - \mu}{\sigma}\right) \qquad (4.01)$$

$$\text{where} \quad \Phi(z) = \frac{1}{\surd(2\pi)} \int_{-\infty}^{z} \exp(-\tfrac{1}{2}t^2)\, dt$$

The values of ℓ which are of interest in the present context are those in the upper tail of the distribution. While there have been various suggestions as to an appropriate value of ℓ to take as a lower limit of "heavy consumption", there is still no universally agreed value. Such a value undoubtedly depends on sex, and possibly on age and various other factors. As representative values we shall take ℓ = 60, 80 and 100 g/day.

We are interested in how $p(\ell)$ varies with changing values of μ and σ. Figure 4.3 shows $p(\ell)$ plotted as a function of μ for constant σ = 1.5, for ℓ = 60, 80 and 100. For each value of ℓ, $p(\ell)$ is approximately quadratic in μ. Figure 4.4 shows that this relationship holds approximately true for changing values of σ. The figure plots p(80) as a function of μ for values of σ = 0.5, 1.0, 1.5 and 2.0. Thus the lines marked ℓ = 80 in Figure 4.3, and σ = 1.5 on Figure 4.4 are the same. This range of values for σ covers the values commonly found in Australian data (see Chapter 6).

In the present study it is often more relevant to consider changes related to the mean consumption, ξ, and so we recast these two figures in terms of ξ rather than μ. These are presented as Figures 4.5 and 4.6. Figure 4.5 shows $p(\ell)$, ℓ = 60, 80, 100, plotted as a function of the mean consumption, ξ, for constant σ = 1.5. Comparison with Figure 4.3 shows that, for mean consumption greater than about 10 g/day, the rate of increase of $p(\ell)$ is approximately linear. To illustrate how the relationship changes with

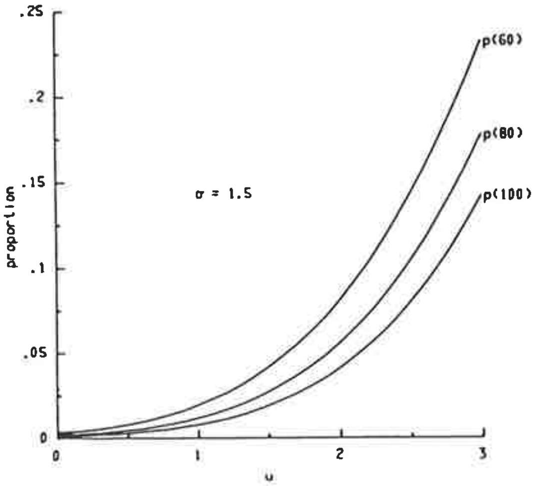

Figure 4.3 The proportion of drinkers consuming in excess of 60, 80 and 100 g alcohol/day, as a function of μ.

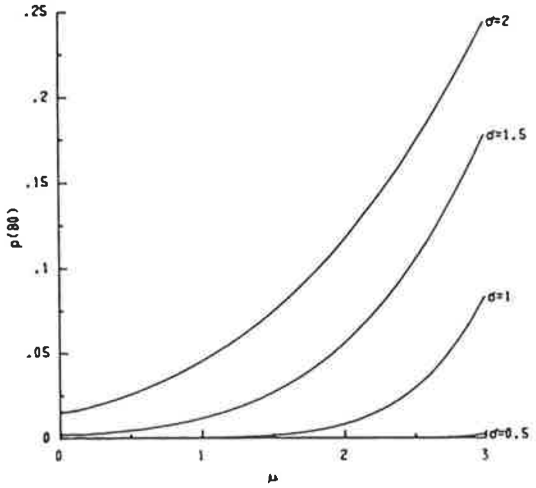

Figure 4.4 The proportion of drinkers consuming in excess of 80 g alcohol/day, as a function of μ and σ.

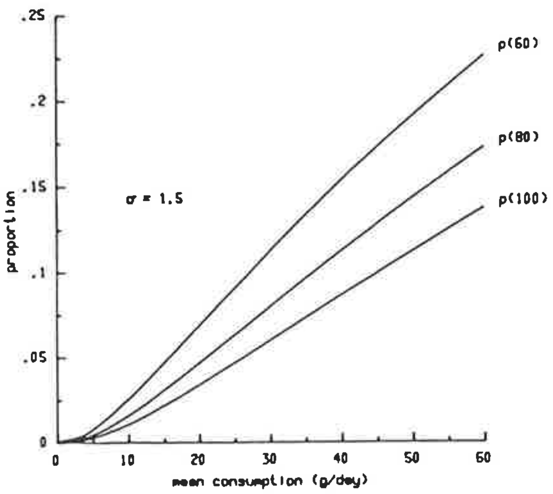

Figure 4.5 The proportion of drinkers consuming in excess of 60, 80 and 100 g alcohol/day, as a function of the mean consumption.

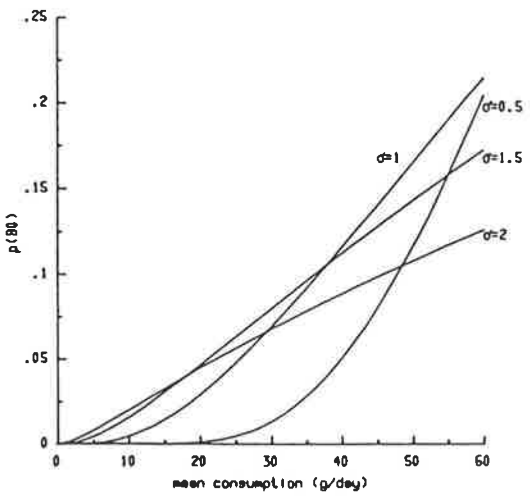

Figure 4.6 The proportion of drinkers consuming in excess of 80 g alcohol/day, as a function of the mean consumption and σ.

varying values of σ, Figure 4.6 shows p(80) plotted as a function of the mean consumption for σ = 0.5, 1.0, 1.5 and 2.0. As before, the lines marked ℓ = 80 on Figure 4.5, and σ = 1.5 on Figure 4.6, are the same. In this instance however, we see that a change in the value of σ can have a considerable effect on the proportion of consumers drinking more than 80 g/day. This is because ξ is a function of both μ and σ.

Consider now a population with a mean consumption of 30 g/day. Figure 4.7 shows how p(80) changes as the mean consumption changes, giving percentage changes in both ordinate and abscissa for the usual range of values of σ. We note that the change in p(80) is sensitive to the value of σ. For instance, a 10% decrease in mean consumption will be accompanied by only a 10% decrease in the number of heavy consumers if σ = 2, but if σ = 1, the decrease will be about 18%. For σ = 0.5, the decrease in p(80) is even more dramatic, and is of the order of 35%.

An alternative way of studying these interdependencies is through a contour map of $p(\ell)$ as a function of either μ or ξ, and σ. For a contour $p(\ell) = p_0$, we have

$$p(\ell) = p_0 = 1 - \Phi\left(\frac{\log \ell - \mu}{\sigma}\right)$$

Therefore

$$\Phi^{-1}(1 - p_0) = \frac{\log \ell - \mu}{\sigma}$$

where Φ^{-1} can be read from tables of the cumulative normal distribution. Thus to draw contours as functions of μ and σ, we can select a range of values for μ, and calculate σ as

$$\sigma = \frac{\log \ell - \mu}{\Phi^{-1}(1 - p_0)} .$$

Figure 4.8 shows such contours for p(80) = 0.01, 0.05, 0.1, 0.2. The

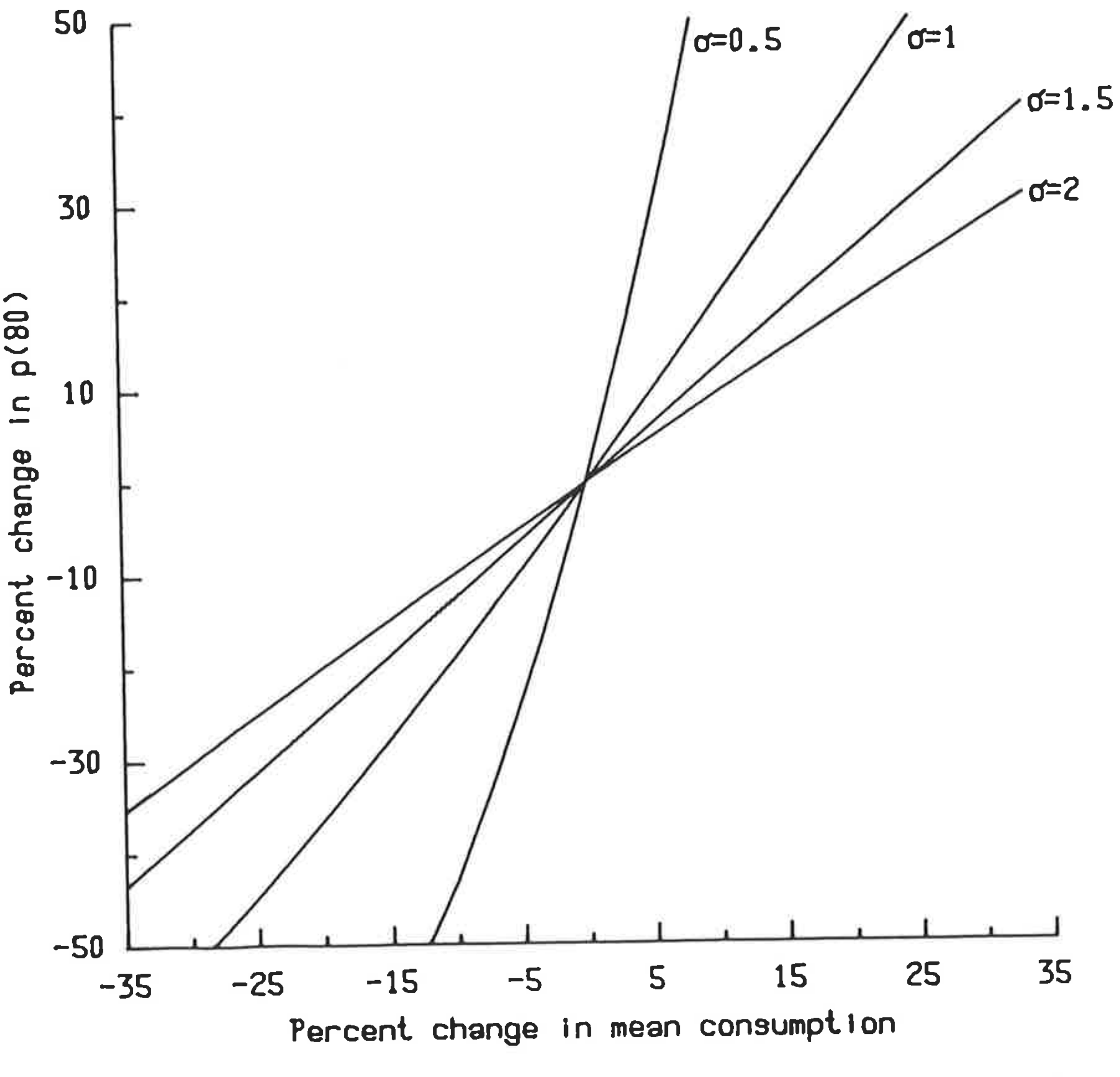

Figure 4.7 Change in percentage of drinkers consuming in excess of 80 g/day as a function of the change in mean consumption from 30 g/day.

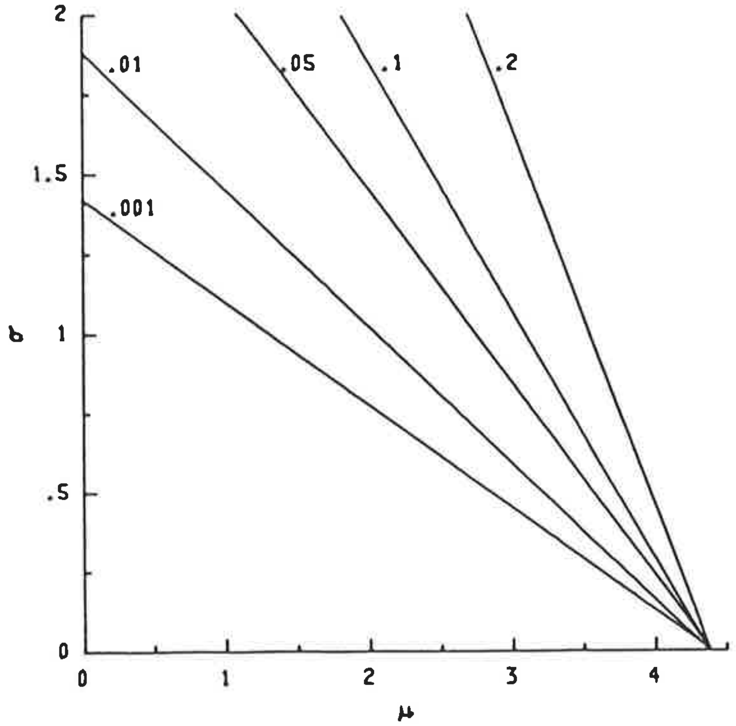

Figure 4.8 Contours of p(80) as a function of μ and σ.

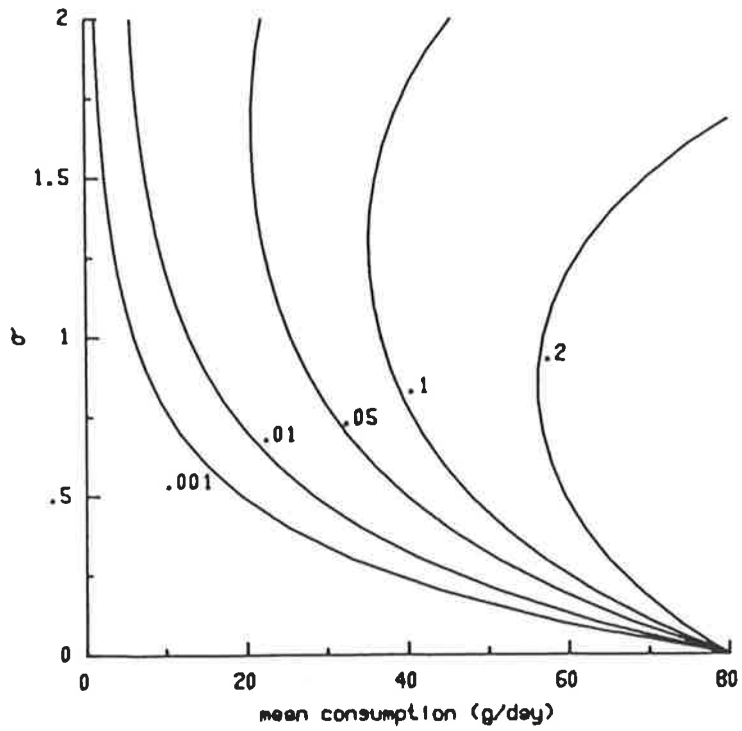

Figure 4.9 Contours of p(80) as a function of the mean consumption and σ.

contours form a fan of lines radiating from the point (μ = log 80, σ = 0). For other values of ℓ, a similar fan will obtain, but will radiate from a different point on the μ axis.

We can express the contours in terms of the mean consumption, ξ, rather than μ, by calculating $\xi = \exp(\mu + \frac{1}{2}\sigma^2)$; Figure 4.9 shows the contours plotted as a function of ξ and σ. The contours now assume a curvilinear form.

4.2 The three parameter lognormal distribution

We introduce a third parameter into the two parameter lognormal distribution considered in section 3.1, such that a simple displacement of X, and not X itself, has a two parameter lognormal distribution. Thus

$$X' = X - \tau \sim \mathrm{LN}\,(\mu,\ \sigma)$$

$$\text{or } X \sim \mathrm{LN}\,(\mu,\ \sigma,\ \tau) \qquad \tau<x<\infty$$

The range of X is now from τ to infinity, with τ being a "threshold" parameter.

The density function is

$$f(x) = \frac{1}{\sigma(x-\tau)\sqrt{(2\pi)}} \exp\left\{ -\frac{1}{2\sigma^2}(\log(x-\tau) - \mu)^2 \right\} dx \qquad \tau < x < \infty$$

Since the 3 parameter distribution is a simple displacement of the two parameter distribution, the location characteristics are each increased by τ. Thus

$$\mathrm{mean}(X) = \tau + \exp(\mu + \tfrac{1}{2}\sigma^2)$$

$$\mathrm{median}(X) = \tau + \exp(\mu)$$

$$\mathrm{mode}(X) = \tau + \exp(\mu - \sigma^2).$$

The variance of X remains unchanged.

Figure 4.10 shows the frequency curves of two three parameter lognormal distributions with $\mu = 2$, $\sigma = 1$. This illustrates the displacement of the two parameter curve.

Equation (4.01) for the proportion of consumers above ℓ becomes

$$p(\ell) = 1 - \Phi\left(\frac{\log(\ell - \tau) - \mu}{\sigma}\right) \qquad \ell > \tau$$

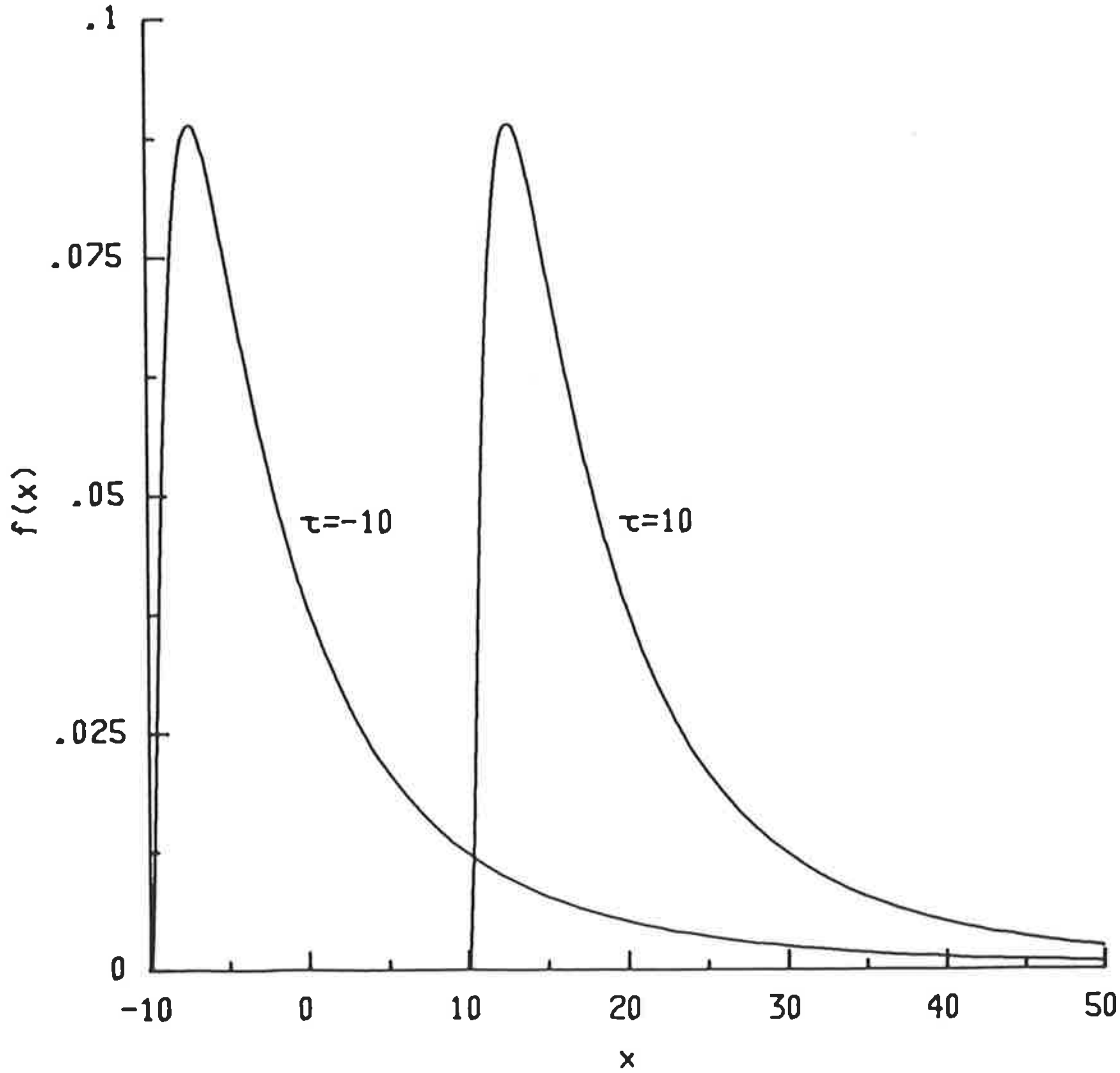

Figure 4.10 Frequency curves for the three parameter lognormal distribution with two values of τ (for μ=2, σ=1).

4.3 Truncated and censored lognormal distributions

We consider below the three parameter lognormal distribution. The results for the two parameter case follow by taking $\tau = 0$.

Suppose we have a variate $X \sim \mathrm{LN}(\mu, \sigma, \tau)$ with that part of the distribution for which $X \leq \zeta$ removed. Then the distribution is said to be <u>truncated</u>, and ζ is termed the <u>point of truncation</u>.

The distribution function of a truncated distribution can be specified as

$$\Pr(X \leq x) = \begin{cases} 0 & 0 < x \leq \zeta \\ \dfrac{\Pr(\zeta < X \leq x)}{\Pr(X > \zeta)} & \zeta < X < \infty \end{cases}$$

The mean of the truncated distribution is given by

$$E(X) = \tau + \exp\left(\mu + \tfrac{1}{2}\sigma^2\right) \frac{\Pr\{X > \zeta \mid X \sim \mathrm{LN}(\mu + \sigma^2, \sigma, \tau)\}}{\Pr\{X > \zeta \mid X \sim \mathrm{LN}(\mu, \sigma, \tau)\}}$$

which was shown by Quensel (1945) for the two parameter case.

In some cases, however, we may have limited knowledge about X in the range $(0, \zeta)$, i.e. we may know the proportion of the distribution lying below ζ, but not the exact values of the variate in this range. The distribution is then said to be <u>censored</u>, and ζ is the <u>point of censorship</u>. The specification of the censored distribution is

$$\Pr(X \leq x) = \begin{cases} \Pr(X \leq \zeta) & 0 < x \leq \zeta \\ \Pr(X \leq x) & \zeta < x < \infty \end{cases}$$

4.4 Estimation of lognormal distributions from grouped data

4.4.1 Introduction Throughout this thesis we shall use the method of maximum likelihood for estimation of the parameters of distributions. We are concerned with estimating from grouped data, since it is in that form that data from alcohol consumption studies are usually presented. In Chapter 7 we will formulate maximum likelihood estimation from grouped data as iterated weighted regression; the present section sets up the necessary details to use that method for the estimation of lognormal distributions.

4.4.2 Maximum likelihood estimation of lognormal distributions from grouped data - a brief review In fitting lognormal distributions to grouped data, we firstly note that such a set of grouped data is equivalent to a sample from a multinomial distribution, with class probabilities determined by the underlying lognormal distribution. For the multinomial distribution, Rao (1957) and Kulldorff (1961) have established sufficient conditions under which the maximum likelihood estimates of the parameters are consistent and asymptotically efficient.

Fisher (1931) and Stevens (in Bliss, 1937) made early contributions to the maximum likelihood solution for censored and truncated normal distributions for continuous data. Since then, several authors have examined maximum likelihood estimation of the normal and lognormal distributions from grouped data, and also at the special problems of truncation and censoring. Most of the literature has been concerned with methods which were tractable for hand calculation: e.g. Gjeddebaek (1949) solved the likelihood equations for the case of the normal distribution with the aid of tables of

$$z_1(x,y) = -\frac{\phi(x+y) - \phi(x)}{\Phi(x+y) - \Phi(x)}$$

and

$$z_2(x,y) = \frac{\phi'(x+y) - \phi'(x)}{\Phi(x+y) - \Phi(x)}$$

Swamy (1960) extended this method to the case of truncated and censored observations.

Grundy (1952) used adjusted moments to find the maximum likelihood estimates. This approach used tables of truncated and censored normal distributions given by Hald (1949).

An alternative method to either of these is the method of scoring for parameters (Fisher, 1935, 1954; see e.g. Bailey, 1961) giving an iterative set of equations. This procedure is a modified Newton-Raphson method of solving equations. The method is advocated by Kulldorff (1961) and by Tallis and Young (1962), and mentioned in Aitchison and Brown (1957) in the context of probit analysis.

Cohen (1951) used an iterative technique to fit the three parameter log-normal, based on the direct solution of the likelihood equations, but advocated an alternative approach based on the least observed value, on the grounds that it was more easily computed. Hill (1963) showed that there exist paths along which the likelihood function for the three parameter log-normal can tend to infinity. However solution of the likelihood equations leads, in most cases, to local maximum likelihood estimates.

Other aspects of the problem have been studied. Gjeddebaek (1956) and Swamy (1962, 1963) have considered the loss of information due to grouping in, respectively, the normal, censored normal, and truncated normal cases. Kale (1964, 1966) has considered this information loss in a more general situation.

The problem of initial parameter estimates for fitting the 3 parameter lognormal by scoring for parameters was considered by Michelini (1972).

4.4.3 Details necessary for maximum likelihood fitting of lognormal distributions using iterated weighted regression

4.4.3.1 Introduction In Chapter 7 we shall formulate maximum likelihood estimation from grouped data as iterated weighted regression. In using the method to fit a lognormal distribution, we will need the first derivatives of the class probabilities with respect to the parameters. We consider both the untruncated and the truncated three parameter lognormal distributions. The results for the two parameter cases are immediately given by ignoring derivatives with respect to τ, and taking $\tau = 0$ in equation (4.02) below. Estimation of the censored distributions presents no new problems since the censored part of the distribution is regarded as forming the first class of the grouped distribution.

4.4.3.2 The untruncated case Let $X \sim \mathrm{LN}(\mu, \sigma, \tau)$ and suppose the data to be grouped with lower class boundaries

$$x_1 , x_2 , \dots , x_m \qquad x_1 = \tau$$

and the frequencies in the class intervals being

$$a_1 , a_2 , \dots , a_m \qquad \sum_{i=1}^{m} a_i = n$$

and relative frequencies

$$f_1 , f_2 , \dots , f_m \qquad f_i = a_i/n , \qquad \sum_{i=1}^{m} f_i = 1.$$

Let

$$y_i = \log(x_i - \tau) \qquad i = 1, \ldots, m \tag{4.02}$$

Then

$$z_i = \frac{y_i - \mu}{\sigma} \sim N(0,1).$$

Let

$$p_i = E[f_i].$$

Then

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

where Φ represents the cumulative standard normal distribution function, and $\Phi(z_1) = 0$, $\Phi(z_{m+1}) = 1$.

For notational simplicity we write $\Phi = \Phi(z)$ for the cumulative normal distribution function, $\phi = \phi(z)$ for the normal frequency function, and define an operator Δ such that

$$\Delta(a) = a_{i+1} - a_i$$

$$\Delta(ab) = a_{i+1}b_{i+1} - a_i b_i \quad \text{etc.}$$

Thus for example

$$p_i = \Delta(\Phi)$$

and

$$\Delta(\phi z^2 e^{-y}) = \phi(z_{i+1})z_{i+1}^2 e^{-y_{i+1}} - \phi(z_i)z_i^2 e^{-y_i}.$$

We have

$$\frac{\partial y}{\partial \tau} = -e^{-y}, \qquad \frac{\partial \Phi}{\partial z} = \phi, \qquad \frac{\partial \phi}{\partial z} = -\phi z$$

and

$$\begin{bmatrix} \partial z/\partial\mu \\ \partial z/\partial\sigma \\ \partial z/\partial\tau \end{bmatrix} = -\frac{1}{\sigma}\begin{bmatrix} 1 \\ z \\ e^{-y} \end{bmatrix}$$

$$\begin{bmatrix} \partial\Phi/\partial\mu \\ \partial\Phi/\partial\sigma \\ \partial\Phi/\partial\tau \end{bmatrix} = -\frac{\phi}{\sigma}\begin{bmatrix} 1 \\ z \\ e^{-y} \end{bmatrix}$$

$$\begin{bmatrix} \partial\phi/\partial\mu \\ \partial\phi/\partial\sigma \\ \partial\phi/\partial\tau \end{bmatrix} = \frac{\phi z}{\sigma}\begin{bmatrix} 1 \\ z \\ e^{-y} \end{bmatrix} .$$

Then

$$p_{i_\mu} = \frac{\partial}{\partial\mu} p_i = \frac{\partial}{\partial\mu}\Delta(\Phi) = -\frac{1}{\sigma}\Delta(\phi)$$
$$p_{i_\sigma} = \frac{\partial}{\partial\sigma} p_i = \frac{\partial}{\partial\sigma}\Delta(\Phi) = -\frac{1}{\sigma}\Delta(\phi z)$$
$$p_{i_\tau} = \frac{\partial}{\partial\tau} p_i = \frac{\partial}{\partial\tau}\Delta(\Phi) = -\frac{1}{\sigma}\Delta(\phi e^{-y})$$

We also record the second derivatives:

$$p_{i_{\mu\mu}} = \frac{\partial}{\partial\mu}\left(-\frac{1}{\sigma}\Delta(\phi) \right) = -\frac{1}{\sigma^2}\Delta(\phi z)$$
$$p_{i_{\mu\sigma}} = \frac{\partial}{\partial\sigma}\left(-\frac{1}{\sigma}\Delta(\phi) \right) = -\frac{1}{\sigma^2}\{ \Delta(\phi z^2) - \Delta(\phi) \}$$
$$p_{i_{\sigma\sigma}} = \frac{\partial}{\partial\sigma}\left(-\frac{1}{\sigma}\Delta(\phi z) \right) = -\frac{1}{\sigma^2}\{ \Delta(\phi z^3) - 2\Delta(\phi z) \}$$
$$p_{i_{\mu\tau}} = \frac{\partial}{\partial\tau}\left(-\frac{1}{\sigma}\Delta(\phi) \right) = -\frac{1}{\sigma^2}\Delta(\phi z e^{-y})$$
$$p_{i_{\sigma\tau}} = \frac{\partial}{\partial\tau}\left(-\frac{1}{\sigma}\Delta(\phi z) \right) = -\frac{1}{\sigma^2}\{ \Delta(\phi z^2 e^{-y}) - \Delta(\phi e^{-y}) \}$$
$$p_{i_{\tau\tau}} = \frac{\partial}{\partial\tau}\left(-\frac{1}{\sigma}\Delta(\phi e^{-y}) \right) = -\frac{1}{\sigma^2}\{ \Delta(\phi z e^{-2y}) + \sigma\Delta(\phi e^{-2y}) \}$$

To obtain the relevant expressions for the two parameter distribution, we set $\tau = 0$ and ignore derivatives with respect to τ.

4.4.3.3 <u>The truncated case</u> Suppose now the lower $t-1$ classes of the grouped distribution are truncated, i.e. the point of truncation is $\zeta = x_t$. Then we have lower class boundaries

$$x_t,\ x_{t+1},\ \ldots\ ,\ x_m$$

with frequencies

$$a_t,\ a_{t+1},\ \dots,\ a_m \qquad \sum_{i=t}^{m} a_i = n$$

and relative frequencies

$$f_t\ ,\ f_{t+1}\ ,\ \dots\ ,\ f_m \qquad f_i = a_i/n \qquad \sum_{i=t}^{m} f_i = 1$$

As before we consider the three parameter case and the two parameter case then follows by ignoring derivatives with respect to τ and setting $\tau = 0$ in equation (4.03) below. We have

$$y_i = \log\,(x_i - \tau) \qquad i = t\ ,\ \dots\ ,\ m \tag{4.03}$$

and

$$z_i = \frac{y_i - \mu}{\sigma} \sim N(0,1).$$

We write

$$E[f_i] = q_i$$

for distinction between the truncated and untruncated cases;

$$q_i = \frac{\Phi(z_{i+1}) - \Phi(z_i)}{1 - \Phi(z_t)}$$

and the loglikelihood function is

$$\log \mathcal{L} = \text{constant} + n \sum_{i=t}^{m} f_i \log q_i.$$

Then we have

$$\frac{\partial \log \mathcal{L}}{\partial \theta_j} = n \sum_{i=t}^{m} \frac{f_i}{q_i}\, q_{i_{\theta_j}}$$

To calculate the derivatives with respect to the parameters we use the same notation as for the untruncated case, with the addition of

$$\Phi_t = \Phi(z_t) \quad \text{and} \quad \phi_t = \phi(z_t)$$

We have

$$q_{i_\mu} = \frac{\partial}{\partial \mu}\left(\frac{\Delta(\Phi)}{1 - \Phi_t} \right) = -\frac{1}{\sigma(1 - \Phi_t)}\, \{\Delta(\phi) + \phi_t q_i\}$$

$$q_{i_\sigma} = \frac{\partial}{\partial\sigma}\left(\frac{\Delta(\Phi)}{1-\Phi_t}\right) = -\frac{1}{\sigma(1-\Phi_t)}\{\Delta(\phi z) + \phi_t z_t q_i\}$$

$$q_{i_\tau} = \frac{\partial}{\partial\tau}\left(\frac{\Delta(\Phi)}{1-\Phi_t}\right) = -\frac{1}{\sigma(1-\Phi_t)}\{\Delta(\phi e^{-y}) + \phi_t e^{-y_t} q_i\}$$

$$q_{i_{\mu\mu}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\Delta(\phi z) + \phi_t z_t q_i + 2\sigma\phi_t q_{i_\mu}\}$$

$$q_{i_{\mu\sigma}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\Delta(\phi z^2) + \phi_t z_t^2 q_i + \sigma\phi_t q_{i_\sigma} + \sigma(1 - \Phi_t + \Phi_t z_t) q_{i_\mu}\}$$

$$q_{i_{\sigma\sigma}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\Delta(\phi z^3) - \Delta(\phi z) - (\Phi_t z_t + \phi_t z_t^2) q_i + \sigma(1 - \Phi_t + \Phi_t z_t + \phi_t z_t) q_{i_\sigma}\}$$

$$q_{i_{\mu\tau}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\Delta(\phi z e^{-y}) + \phi_t z_t e^{-y_t} q_i + \sigma\phi_t q_{i_\tau} + \sigma\phi_t e^{-y_t} q_{i_\mu}\}$$

$$q_{i_{\sigma\tau}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\Delta(\phi z^2 e^{-y}) - \Delta(\phi e^{-y}) - (\phi_t e^{-y} - \phi_t z_t^2 e^{-y}) q_i + \sigma\phi_t z_t q_{i_\tau} - \sigma\phi_t z_t e^{-y_t} q_{i_\sigma}\}$$

$$q_{i_{\tau\tau}} = -\frac{1}{\sigma^2(1-\Phi_t)}\{\sigma\Delta(\phi e^{-2y}) + \Delta(\phi z e^{-2y}) + (\sigma\phi_t e^{-2y_t} + \phi_t z_t e^{-2y_t}) q_i + 2\sigma\phi_t e^{-y_t} q_{i_\tau}\}$$

We note that the above expressions can be written in terms of the probabilities of the untruncated case, e.g.

$$q_{i_\mu} = \frac{1}{(1-\Phi_t)}\left(p_{i_\mu} - \frac{\phi_t}{\sigma} q_i\right)$$

but the expressions given above are more convenient for computation since they do not require the prior fitting of the untruncated distribution.

4.5 The gamma distribution

4.5.1 Definition Let X = mean individual alcohol consumption, $0 < X < \infty$. Then X is said to follow a (two parameter) gamma distribution if

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\,dx \qquad \alpha > 0;\ \beta > 0$$

α is a shape parameter, and β is a scale parameter.

We note that if $\alpha = 1$ the distribution reduces to an exponential distribution with parameter $1/\beta$.

4.5.2 Characteristics The rth moment about the origin is given by

$$\begin{aligned} \mu_r' &= \int_0^{\infty} \frac{x^r x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\,dx \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \frac{x^{\alpha+r+1} e^{-x/\beta}}{\beta^{\alpha}}\,dx \\ &= \frac{\beta^r}{\Gamma(\alpha)} \int_0^{\infty} (\frac{x}{\beta})^{\alpha+r+1} e^{-x/\beta}\,\frac{dx}{\beta} \\ &= \frac{\beta^r\Gamma(\alpha+r)}{\Gamma(\alpha)} \end{aligned}$$

Hence

$$\begin{aligned} \text{Mean } (X) = E[X] = \mu_1' &= \beta\,\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\ &= \alpha\beta \end{aligned}$$

since

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha).$$

$$\begin{aligned} \text{Var } (X) &= \mu_2' - (\mu_1')^2 \\ &= \frac{\beta^2\ \Gamma(\alpha+2)}{\Gamma(\alpha)} - \alpha^2\beta^2 \\ &= \alpha(\alpha+1)\beta^2 - \alpha\beta^2 \\ &= \alpha\beta^2\ . \end{aligned}$$

The gamma distribution has a single mode at $\beta(\alpha-1)$, provided $\alpha \geq 1$. If $\alpha < 1$, the distribution is asymptotic to the y axis.

In the present study, values of α commonly lie in the range 0.2 to 2.0, while β values up to 50 are common, although some are much larger. Figure 4.11 gives frequency curves for $\alpha = 1$ and $\beta = 10, 20, 30$. Figure 4.12 gives frequency curves for $\beta = 20$ and $\alpha = 0.5, 1, 2$.

4.5.3 <u>The proportion of heavy drinkers</u> Suppose X has a gamma distribution with parameters α and β.

i.e. $$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} dx$$

Putting $\alpha = \nu/2$, $x = \chi^2\beta/2$ gives

$$f(\chi^2) = \frac{(\chi^2)^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}\chi^2}}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)} d\chi^2$$

which is the frequency function for a chisquare variable with ν degrees of freedom.

We can use this fact and tables of the chisquare distribution function to investigate the behaviour of the upper tail of the gamma distribution with changes in α and β.

Thus

$$p(\ell) = \Pr(X > \ell)$$
$$= \Pr(\chi^2_{2\alpha} > 2\ell/\beta).$$

Figure 4.13 shows $p(\ell)$ plotted as a function of α for constant $\beta = 20$, and $\ell = 60, 80, 100$. The mean consumption ($= \alpha\beta$) is also shown on the figure. The curves have the same general shape as the equivalent ones for the log-normal distribution, shown in Figure 4.3.

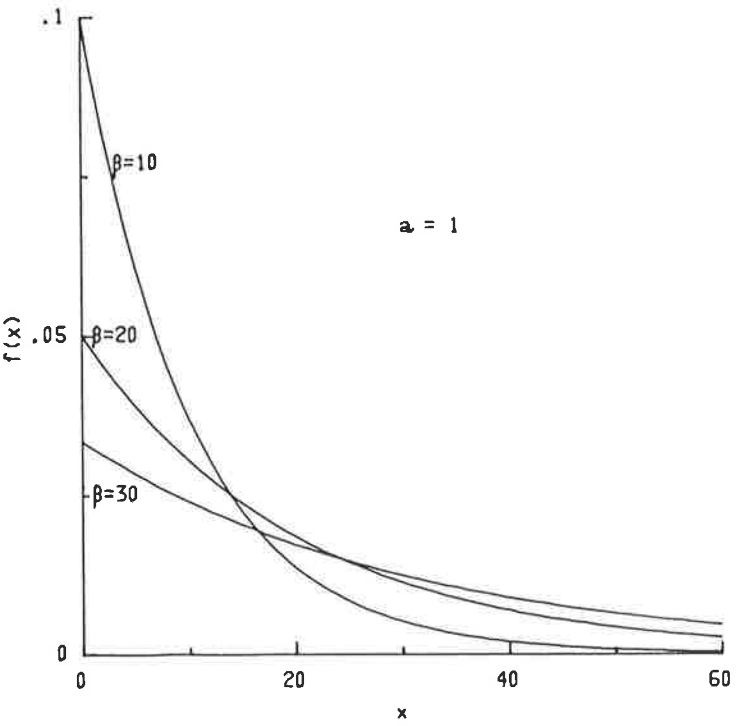

Figure 4.11 Frequency curves of the gamma distribution with $\alpha=1$ and three values of β.

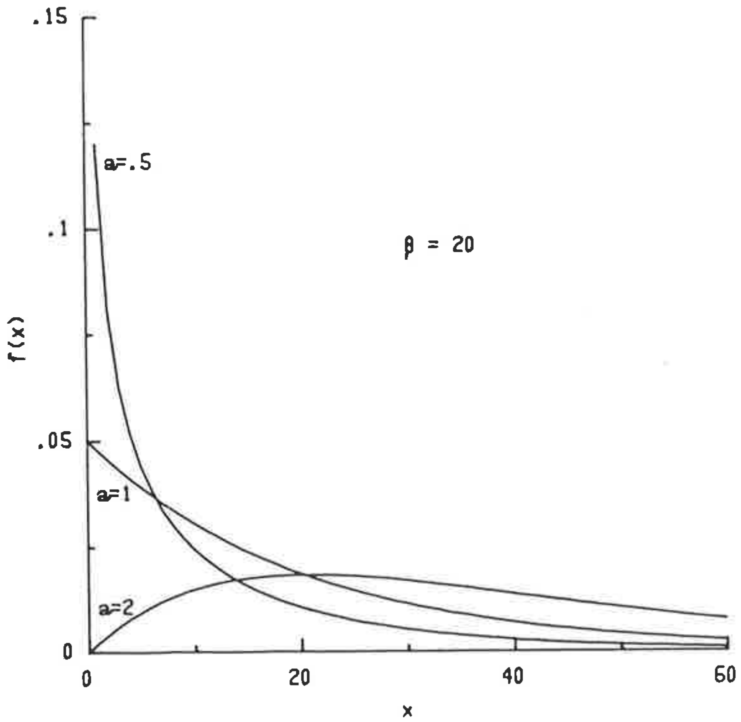

Figure 4.12 Frequency curves of the gamma distribution with $\beta=20$ and three values of α.

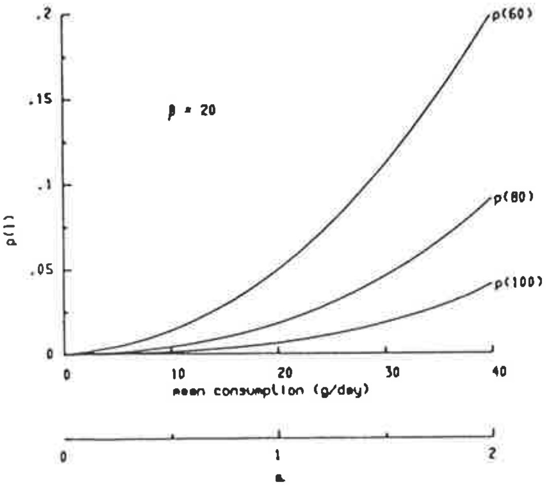

Figure 4.13 The proportion of drinkers consuming in excess of 60, 80 and 100 g alcohol/day, as a function of α and the mean consumption.

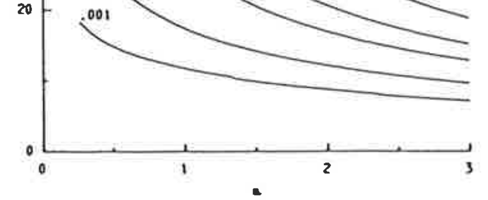

Figure 4.14 Contours of p(80) as functions of α and β.

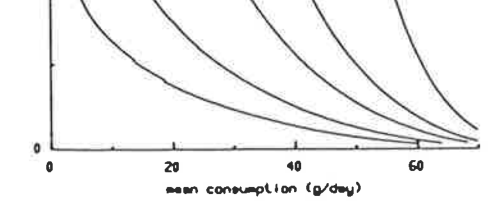

Figure 4.15 Contours of p(80) as functions of β and the mean consumption.

Figure 4.14 shows contours of p(80) as a function of α and β while Figure 4.15 shows the contours as functions of the mean consumption and β. The similarity of Figures 4.8 and 4.9 respectively from the lognormal distribution is apparent.

4.6 A model relating age subpopulations

The models in this chapter have assumed that we are dealing with one homogeneous population, and can fit a single distribution to the consumption data from a sample of that population. In practice, the consumption of an entire population, such as the residents of Australia, or even of one city, will not be homogeneous, but there may be differences in consumption patterns with differences in sex, age groups, ethnic background, climate, beverage type, and so on.

In published consumption data, the most common variables which are used to stratify the data are sex and age. While there are physiological and social reasons why we may wish to consider a separate model for each sex, it is useful to consider a model which will, for a given sex, describe the consumption pattern over all age groups. Empirical evidence (see Chapter 6) suggests that in many cases, the same form of model may be appropriate for all ages, but with changing parameters. For notational simplicity, we consider a two parameter model, but the generalisation beyond two parameters is obvious.

Suppose we have r ages or age groups, which we represent as t_i, $i = 1, \ldots, r$. At age t_i we have $X(t_i)$ = mean individual alcohol consumption. We assume that the one functional form of model, depending on parameters θ_1 and θ_2, provides a satisfactory description of the consumption data at each age group, and then we assume further that the parameters θ_1 and θ_2 are functions of age. That is

$$X(t_i) = f[\theta_1(t_i), \theta_2(t_i)], \qquad i = 1, \ldots, r$$

In the context of this study, the most common form of $f[]$ will be a lognormal or gamma distribution.

The form of the functions θ_1 and θ_2 is in practice limited by the number of consumption groups available from the data. If we assume a quadratic relation with age, we have

$$\theta_1(t_i) = a_0 + a_1 t_i + a_2 t_i^2$$

and

$$\theta_2(t_i) = b_0 + b_1 t_i + b_2 t_i^2$$

In fitting the model, we estimate the six parameters a_0, a_1, a_2, b_0, b_1, b_2, and can then calculate the two parameters θ_1 and θ_2 for any given age t_i. In practice, as with any regression procedure, great care should be taken if extrapolating outside the range of the ages t_i, but the procedure provides a useful means of interpolating within that range.

Chapter 5

Australian data on the distribution of alcohol consumption.

5.1 Methods of measuring individual alcohol consumption

5.1.1 Introduction The reasons for measuring alcohol consumption are varied and include

a. the estimation in a population of overall consumption, its distribution and trend with time,

b. characterization of the drinking population, or groups within it who are consuming at a high risk level,

c. the evaluation of intervention programmes aimed at changing alcohol consumption patterns, and

d. the linking of alcohol consumption with other characteristics of a population.

The variables of interest will vary with the reason for study. Variables which have commonly been studied are the type of beverage, quantity and frequency of intake, the social circumstances, past history and short and long term patterns of consumption. In estimating the distribution of alcohol consumption we are mainly interested in the amount of absolute alcohol consumed by an individual in a specified time period.

Various methods have been used to measure such consumption and there is no universally agreed "best" method. A convenient classification of methods is provided by considering methods of elucidating present or past consumption.

5.1.2 <u>Present consumption</u> Present consumption can be measured by either *self recording* or *direct observation*. The methods have been little used in alcohol studies; self recording can suffer from the disadvantage of tending to modify the usual intake (Baghurst, 1978), although Sudman (1980) suggests that this is not a problem if several items are reported simultaneously. Plant *et al* (1977) described a reliable method for assessing alcohol consumption in public bars by direct observation, and suggested that it might be a useful way of investigating the distribution of alcohol consumption in a community where it is believed that most of the drinking is done in public. But this is a doubtful assumption in Australia. The Australian Associated Brewers (AAB) (1978) have noted that in Australia, between 1967–68 and 1976–77 there was a trend towards off-licence consumption with sales of packaged beer increasing from 43% to 57% of total beer production over the period. In 1977 alcohol consumed as beer accounted for 68% of the apparent alcohol consumption (AAB, 1978).

Kamien (1975a, 1975b) however has used participant observation as a method of observing drinking in a population of aborigines living in Bourke, N.S.W. A comparison with estimated weekly expenditure on alcohol showed his consumption figures to underestimate consumption by 33% (Kamien, 1978) but there is no way of knowing which figure (consumption or expenditure) is more in error.

A different approach to the measurement of present consumption is the use of physiological tests, such as the sweat patch (Phillips, 1982, 1984). The test is based on the fact that the concentration of ethanol in sweat varies with the amount of alcohol consumed. A small watertight adhesive patch is worn on the skin for about a week, and collects sweat at a steady rate. At the end of the period the patch can be rapidly assayed to give a measurement of the subject's alcohol intake over the period. While there appear to be some difficulties in field use (Phillips *et al*, 1984) the method is a promising way of obtaining accurate information in a non-invasive fashion.

5.1.3 Past consumption A recall of past consumption is the most commonly used method of measuring alcohol intake. Within this category indirect observation, interview and questionnaire techniques are mainly used.

Indirect observation, such as household expenditure on alcohol, does not seem to have been much used in Australia. Estimates of apparent consumption, based on production, sales, imports and exports of alcohol, while giving more accurate information than survey data do not give any idea of the distribution of consumption across the population, although they can serve as a useful check for total reported consumption obtained from serveys.

A different approach to indirect observation is the informant method. This entails having selected individuals report on the drinking practices of groups familiar to them rather than having individuals report on their own drinking behaviour. Smart and Liban (1982) reported that the method yielded higher rates of drinking and of heavy consumption than did estimates based on standard household survey methods.

Both *interview* and *questionnaire* methods can measure past consumption by asking for either actual consumption over some recent period (e.g. 24 hours, seven days, a month) or for "usual" consumption.

a. <u>Actual consumption.</u> In Australia the most commonly used method of recording actual consumption has been the seven day recall using a structured questionnaire administered by interview (Rankin and Wilkinson, 1971; Australian Bureau of Statistics (ABS), 1978) or self-administered (Barwon Regional Association for Alcohol and Drug Dependence, 1977; Baghurst and McMichael, 1978). For each of the last seven days, respondents are asked the amount and type of their consumption; these quantities are then converted to grammes and an average daily intake calculated. In pilot studies for the 1977 ABS survey, Millwood and McKay (1978) considered two alternative questionnaire designs and found a 20% increase in reported daily consumption by first asking the respondents whether or not they had been drinking on each of the last seven days, and then asking for details of each drinking occasion, rather than immediately asking for details when a drinking occasion was given. This confirmed results previously observed in Scotland by Dight (1976).

Australian users of the seven day recall have also made provision for the respondents to state whether or not last week's intake was a typical one, and if not, to give a typical week's intake. However the uses made of this information vary. The ABS used the question to relieve any tension in the interview resulting from respondents perhaps feeling that they had said too much, or that their drinking behaviour might be seen as excessive (Millwood and McKay, 1978). Baghurst and McMichael (1978) used the typical week's consumption as their reported consumption, using the question on actual consumption as a means of getting the respondent used to the question format (pers. comm.) They reported finding no difference between the consumption of those who reported last week as typical and those who did not. This conflicts with the findings of Chick *et al* (1981) who, using blood tests to corroborate reported consumption of Scottish drinkers, found that those who claimed last week's consumption was atypical had heavier consumption than those who did not. On the other hand, they discovered only a trivial difference between their last week's and their typical week's consumption, and concluded that they appeared to be attempting to deny habitual heavy consumption. Millwood and McKay (1978) and Dight (1976) concluded that the error in measures of "last week" drinking is far preferable to the even larger respondent biases present in reports of "usual" alcohol consumption.

b. Usual consumption. Estimation of usual consumption has been used in many Australian surveys (for example, Krupinski *et al*, 1967; Encel *et al*, 1972; Selge, 1975; McCall *et al*, 1978; Egger *et al*, 1978). If medical interviews are available, the information can be gleaned in a history-taking situation. Thus Krupinski *et al* (1967) used fifth year medical students as interviewers in a community health survey of Heyfield, Victoria. A somewhat different interviewing technique called "grogcount" has been suggested by O'Neill (1977), particularly for use with excessive consumers; at no stage during the assessment of alcohol intake is it suggested that the client is using abnormally or excessively. Use of non-medical interviewers does not allow these approaches, and a structured interview using a questionnaire has been the most frequently used approach. Typically questions are asked relating to the type of beverage drunk, the usual frequency with which each is drunk, and the quantity consumed on a typical drinking occasion.

Various methods have been proposed for constructing measures of aggregate volume of intake from survey data. Straus and Bacon (1953) first suggested a quantity-frequency (QF) measure. Each individual is placed in one of several qualitative quantity-frequency classes by considering both the usual quantity of alcohol in any form consumed on a drinking occasion, and the frequency of drinking. This index has been modified by e.g. Maxwell (1952), Mulford and Miller (1960) Mulford (1964), Knupfer and Room (1964), to subdivide some of the classes further.

Various derivatives of the QF index have been proposed, among them being

- the AA (absolute alcohol) index (Jessor *et al*, 1968) which provides quantitative levels of intake;
- the QFV (quantity-frequency-variability) index (Cahalan *et al*, 1967) which classifies respondents on a five point scale;
- the QV (quantity-variability) index (Cahalan *et al*, 1969) which classifies respondents into eight categories;
- the VP (volume-pattern) index (Bowman *et al*, 1975) avoids discrete classifications, but requires highly detailed information, implying long interviews;
- the AAQP (absolute alcohol-quantity pattern) index (Little *et al*, 1977);
- the KAT (Khavari alcohol test) index (Khavari and Farber, 1978) which provides total annual alcohol intake;
- the QFA (quantity-frequency, adjusted) index (Armor and Pollich, 1982) which combines the QF and QV indices.

Baghurst (1978) has given a useful review of some of these indices, and notes that most of them suffer from the disadvantage categorising drinkers on the basis of subjective or arbitrary decisions. She says: "A typology which governs the whole subsequent analysis is ... set up often without adequate questioning of underlying assumptions".

5.2 Units of measurement of alcohol consumption

Absolute alcohol intake is usually expressed as grammes per day (g/day), centilitres per day (cl/day), centilitres per week (cl/wk) or litres/year (l/yr). Table 5.1 shows conversion factors between the various units. It has been calculated assuming the specific gravity of alcohol is 0.78945 (at 20°C), 1 week = 7 days, and 1 year = 365.25 days. Figure 5.1 presents a comparison of the units over their typical range.

Table 5.1

Conversion factors between commonly used units of alcohol consumption

(The entries in any row represent equivalent quantities of alcohol)

g/day	cl/day	cl/week	l/yr
1	0.127	0.887	0.463
7.895	1	7	3.653
1.128	0.143	1	0.522
2.161	0.274	1.916	1

The most common unit in which details of alcohol consumption is initially recorded in surveys is glasses of beverage. To remove the effects of differing alcoholic content of the differing beverages, this is usually converted to e.g. grammes of absolute alcohol. This calculation requires knowledge of both the size of glasses used and the alcohol content of the various beverages.

Table 5.2 showing alcohol content of typical Australian drinks, is reproduced from the Australian Associated Brewers (1978).

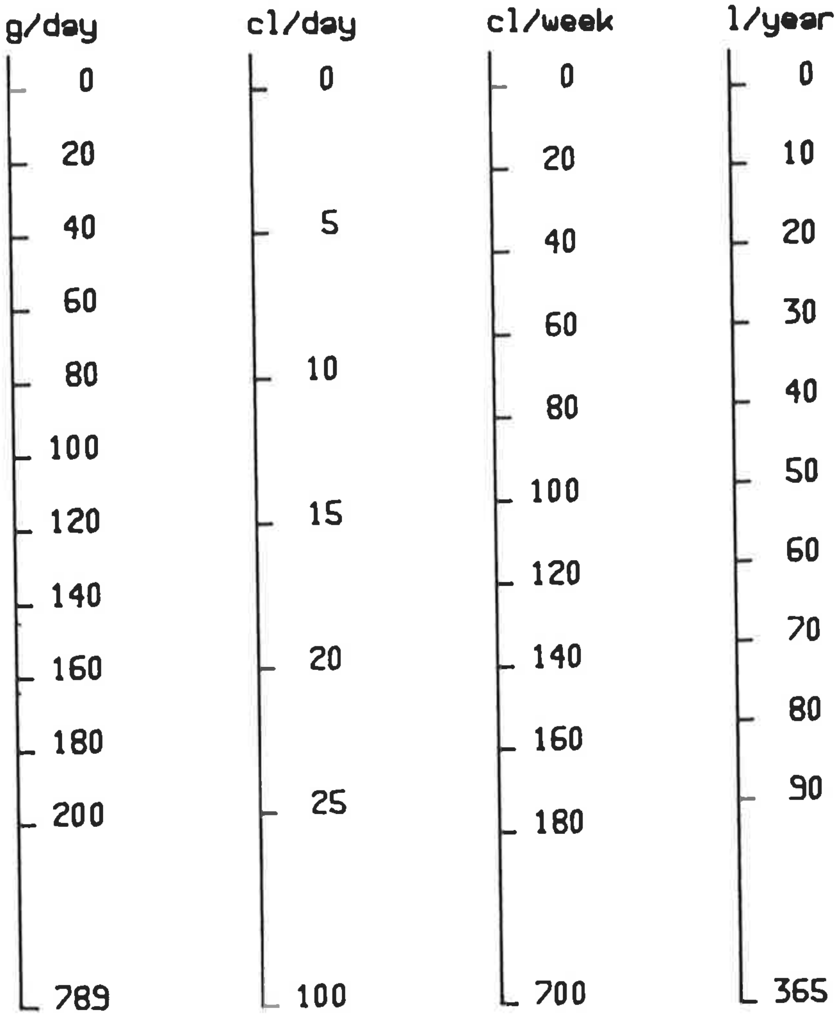

Figure 5.1 A comparison of commonly used units of alcohol consumption.

Table 5.2

Alcohol content of typical Australian drinks

Beverage	alcohol content (% v/v)	"standard drink" volume of beverage (ml)	alcohol content (4) ml	gram
Beer	4.8 (1)	200	9.6	7.6
Wine	11.5 (2)	90	10.4	8.2
Fortified wine	18.5 (2)	60	11.1	8.8
Spirits	38.5 (3)	30	11.6	9.2

Notes:
(1) Australian average
(2) Approximate Australian average
(3) Approximate average for most commonly-encountered beverages. Legal minimum in most states is 37.0
(4) Assumes specific gravity of alcohol = 0.79

As an example of the variation which can exist within these categories, Table 5.3 gives details of alcohol content analyses of South Australian beers in January 1977. (source: SA Brewing Co.)

The alcohol content of standard bottles is given in Table 5.4, using the same assumptions as for Table 5.2.

Regrettably there is no Australian standard for beer glass sizes, and the names given to glass size (middy, 10oz, schooner, etc.) vary between states. Table 5.5 gives the beer glass sizes and their common names used in the various Australian states. The information was provided by the Australian Hotels Association in Queensland, Victoria, Tasmania, South Australia and Western Australia, and in New South Wales.

Table 5.3

Mean values for analyses of SA beers - January 1977

Beer	alcohol content (% v/v)
SA Brewing Co. Ltd.	
Bitter beer (bulk)	4.6
Southwark & West End Bitter (packaged)	4.5
West End Draught (packaged)	4.5
Southwark Premium	4.5
Southwark Pilsener	3.2
Guinness Export Stout	7.3
Cooper & Sons Ltd	
Big Barrel Lager	4.6
Gold Crown Beer	4.8
Diet Beer	4.5
Sparkling Ale	6.2
Light Dinner Ale	5.1
Extra Stout	6.4

Table 5.4

Alcohol content of standard bottle sizes of typical Australian drinks

Beverage	volume (ml)	alcohol (%v/v)	content (g)
Beer - large bottle	750	4.8	28.4
Beer - small bottle	375	4.8	14.2
Wine	750	11.5	68.1
Fortified wine	750	18.5	109.6
Spirits	750	38.5	228.1

Table 5.5

Standard beer glass sizes and their common names - Australia

ml	glass contents approx fl. oz	g alcohol	Qld	NSW ACT	Vic	Tas	SA	WA	NT
115	4	4.4				4 oz	pony		
140	5	5.3	5 oz	5 oz	5 oz			pony	5 oz
170	6	6.4				6 oz	butcher		
200	7	7.6	7 oz	7 oz	glass			glass	
225	8	8.5	8 oz			8 oz			handle
255	9	9.7					schooner		
285	10	10.8	pot	middy	pot	10 oz		middy	
425	15	16.1		schooner			pint	schooner	pint
570	20	21.6		pint					
575	20	21.8						pot	

The most commonly used glasses sizes are as follows (sources as for Table 5.5):

Qld: country 200 ml, city 225 ml

NSW/ACT: 285 ml

Vic: 200 ml, but is being replaced by 285 ml in some parts of the metropolitan area

Tas: 170 ml or 225 ml

SA: 225 ml

WA: 200 ml

NT: 225 ml

5.3 The validity of survey data on alcohol consumption

The main Australian study of validity of survey data on alcohol consumption has been carried out by Millwood and McKay (1978), in conjunction with the ABS survey in February 1977. The survey covered 15947 persons aged 18 years or over in all states, using a seven day recall questionnaire administered by trained interviewers, as outlined in section 5.1.3 (a).

Millwood and McKay found that overall, reported consumption accounted for only 41% of apparent consumption based on production, sales, imports and exports of alcohol (ABS, 1974-75). Fortified wines and spirits were subject to the largest understatement, reported consumption accounting for only 24% and 33% respectively of apparent consumption. Equivalent figures for beer and wine were 44% and 43% respectively. This is in general agreement with similar overseas surveys (Pernanen, 1974; Schmidt, 1973; Midanik, 1982). Wilson (1981) and Popham and Schmidt (1981) have suggested that the understatement is greater for higher consumption categories.

We can examine possible reasons for this understatement under several headings.

a. <u>Incomplete coverage.</u> The ABS survey excluded certain people from the sample: those below eighteen years of age, members of the permanent armed forces, certain diplomatic personnel customarily excluded from census and estimated populations, patients in hospitals and sanitoriums, and inmates of gaols, reformatories. The effect of those omissions is unknown. The sample was based on private and non-private dwellings, and certain groups of moderate to heavy drinkers such as homeless men would not have been included. If these groups drink predominantly one type of beverage, this would help explain the differences in the ratios of reported to apparent consumption for the different beverages.

b. Nonresponse. Millwood and McKay noted that those respondents who were difficult to contact (i.e. required 4-6+ calls to dwelling) had a higher average consumption than those who were relatively easy to locate (1-3 calls). In a Swedish survey on alcohol use (Nilsson and Svensson, 1971), it was discovered that nonrespondents were approximately 3 times more likely to have been registered for drunkenness offences than the respondents.

c. Forgetting. Examination of frequency of recall of "drinking days" showed a marked decline (52% to 47%) in the average proportion of drinkers who reported drinking one day ago to two days ago. Over the seven day recall period, the decline appears approximately exponential. A similar decline appears when considering daily alcohol consumption. This confirms a finding of Pernanen (1974) who also noted that, for Finnish drinkers, frequent drinkers forgot their drinking occasions at a more pronounced rate than infrequent drinkers. Millwood and McKay suggest that yesterday's reported consumption might provide a more accurate estimate of alcohol consumption but they do not carry out the calculations. A shorter recall period would increase the sampling error of estimates, but the gain may be worthwhile.

In the light of this evidence of decreasing recall with time and given the fact that daily consumption increases considerably over the weekend (Millwood and McKay, 1978), the day of interview provides another source of bias if, as was the case with the ABS survey, the day of interview is not balanced for days of the week.

d. Selective reporting. Selective reporting of consumption may not be uniform across a sample, but may vary with sex, age, consumption level or type of alcohol consumed. Popham and Schmidt (1981) gave evidence that under-reporting is greater among heavy users in Canada by comparing the distribution of alcohol purchases as reported in a survey with the distribution from alcohol buying records. Miller *et al* (1977) reported differences in the degree of self-disclosure (Cozby, 1973) for different categories of drinkers. For the abstainer up to moderate drinker category, self-disclosure increases with consumption, but decreases for the heavy drinker category.

e. Interviewer and questionnaire effects. It is known that there are effects of interviewer and questionnaire wording. The questionnaire used in the ABS survey was outlined in section 5.1.3 (a). Blair *et al* (1977) have shown that threatening questions requiring quantified answers are best asked in long questions using wording with which the respondent feels comfortable, allowing the respondent to nominate their own quantity rather than forcing a choice between a number of categories. Plant and Miller (1977) found no overall benefit in disguising questions on drinking behaviour as a health and leisure investigation, but noted that the disguised questionnaire produced a significantly higher mean reported alcohol consumption than the undisguised questionnaire in a working class area, while the reverse occurred in a middle class area. Kirsch *et al* (1965) demonstrated that the most accurate information was obtained using male, non-abstainer, trained and supervised interviewers collecting data in the household of the respondent, and using a structured questionnaire that requires the responses to be made in a set form, predetermined according to the nature of the population being investigated.

McCall *et al* (1978), in a comparison of the 1966 Busselton survey with later ones, stated that it seems probable that heavy drinkers were reluctant to identify themselves in a history taking situation, but most self-administered questionnaires overcame the problem.

There has been interest recently in using randomised response techniques (Volicer and Volicer, 1982) and scrambled randomised response techniques (Eichhorn and Hayre, 1983) to overcome some of the problems involved in asking sensitive questions about alcohol consumption.

In summary, the consequences of the under-reporting are that we are likely to have various biases in the reported distribution of consumption. Incomplete sampling frames, greater nonresponse and selective response of heavy consumers will probably lead to the proportion of heavy consumers being under-represented in the sample. In other countries (e.g. Canada, Finland) methods such as retail sales records provide alternative avenues for study of the distribution, but in Australia these are not available and we must make the best use we can of what data is available.

5.4 The data

Table 5.6 lists forty-seven Australian surveys which contain information, either quantitative or qualitative, on the distribution of alcohol consumption. For each survey the following information is provided:

- the date of the survey
- the population sampled
- reference
- sample size
- the survey method used
- the method used to calculate alcohol consumption
- a description of relevant data available from the survey

Tables 5.7 to 5.27 list the data sets for twenty-one of the surveys which provide some quantitative information about the distribution of alcohol consumption. A breakdown by sex and age is given where possible. Notes following each table give relevant details of construction of scales, etc.

Table 5.6

Australian surveys containing information on the distribution of alcohol consumption

	Date	Population	Reference	Sample size	Survey method	Method of estimation of alcohol consumption	Relevant data available
1	Jan 1965	Heyfield, 200km east of Melbourne	Krupinski *et al* (1967, 1970); Krupinski (1978)	1940	interviews of entire population by 5th year medical students, using history-taking approach	usual drinking habits: type, frequency and quantity	glasses of beer per week for male and female adolescents, adults and elderly [see Table 5.7]
2	1965-66	Perth male "social drinkers"	Tofler *et al* (1969); Tofler and Woodings (1981)	359	interview	not stated	consumption (l. beer or equivalent/day) [see Table 5.8]
3	Jul 1966-Jun 1967	Patients presenting at the Alcoholism Clinic at St. Vincent's Hospital, Melbourne	Wilkinson *et al* (1969)	220	interview by physician or social worker	usual drinking habits: type, frequency and quantity	histogram of daily consumption in grammes, by sex [see Table 5.9]
4	Nov-Dec 1966	Busselton, 240km south of Perth	Curnow *et al* (1969); McCall *et al* (1978)	3393	interview questionnaire of entire population	self-perception as non-, ex-, mild, moderate or heavy drinker	frequency in each category by age and sex separately
5	1967	Australian-born students aged 17-25 at 3 Sydney Universities	Sargent (1979)	2345	self administered questionnaire	not stated	consumption in qualitative groups, by sex
6	Mar-Jun 1968	Prahran, an inner Melbourne suburb	Rankin and Wilkinson (1971)	2163	interview by 5th year medical students	7 day recall, plus usual pattern if last week atypical	daily consumption in grammes by sex and: age, social strata, country of birth; relationship of husband's and wife's consumption; daily consumption by cigarettes smoked [see Table 5.10]
7	1968-69	Sydney metropolitan area residents, 15+ years	Encel and Kotowicz (1970); Encel *et al* (1972)	823	combination of self administered and interviewer administered questionnaire	usual drinking habits: type, frequency and quantity	QFV index for consumption by sex and: age, education, occupation, income, social class, migrant status, religion
8	1969	Busselton, WA	McCall *et al* (1978)	3679	self administered questionnaire to entire population	self-perception as non-, ex-, mild, moderate or heavy drinker	frequencies in each category by sex
9	Jul-Aug 1971	residents aged 14-65 of Manly, a Sydney suburb	George (1972, 1973)	639	self administered questionnaire	usual drinking habits: frequency and quantity	separate tables for frequency x sex x age and quantity x sex
10	Jul-Oct 1971	Canberra adults, 19+ years	Hennessy *et al* (1973)	864	interview (using questionnaire?) by social health visitors, experienced part-time interviewers and postgraduate sociology students	not stated	frequency (never/special occasions/weekends/daily) by sex, with subdivision of daily category into 3 quantities
11	1972	Melbourne (a) secondary students (b) tertiary students (c) working youths	Graves (1973)	a:2042 b:1601 c:307	self administered questionnaire	usual drinking pattern, type and quantity	frequency; consumption (in drinks) on a drinking day
12	Jul-Dec 1972	Aborigines at Bourke, NSW	Kamien (1975b, 1978)	412	participant observation	observation of drinking habits	consumption in g/day by age and sex [see Table 5.11]

Table 5.6

Australian surveys containing information on the distribution of alcohol consumption (continued)

	Date	Population	Reference	Sample size	Survey method	Method of estimation of alcohol consumption	Relevant data available
13	Nov-Dec 1972	Busselton, WA	Cullen and Woodings (1975); McCall *et al* (1978)	3885	self administered questionnaire to entire population	self perception as non-, ex-, mild, moderate or heavy drinker	frequency in each category by age and sex
14	1972-73	South East of SA	Seige (1975)	1027	questionnaire administered by registered public health nurses	usual drinking habits: frequency and quantity (in grammes)	frequency x quantity; frequency by socioeconomic group; amount by socioeconomic group [see Table 5.12]
15	1973	NSW (a) high school students (b) technical college students (c) nurses (d) prisoners (e) probationers (f) delinquents	Bell *et al* (1975)	a:5214 b:1130 c:748 d:188 e:153 f:214	self administered questionnaire	usual frequency; amount on drinking days	frequency for each group
16	1973	Redcliffe, 35km south of Brisbane	Schact *et al* (1976)	994	questionnaire administered by registered nurses	usual frequency and quantity; separately for beer, wines, spirits	consumption (abstainer/moderate/heavy) by sex and age
17	Apr-Jul 1973	Residents of an outer western Sydney suburb, 14-65 years	George (1974)	1011	questionnaire administered by trained interviewer	amount and frequency of use	frequency of use by age and sex
18	1973	Canberra high school students, 11-19 years	Irwin (1976)	4952	self administered questionnaire	self perception as non-, light, medium or heavy user of alcohol	self perceived use by sex and form
19	1974	Canberra high school students, 11-19 years, including matched sample from 1973 survey	Irwin (1976)	5138 (match 2612)	self administered questionnaire	self perception as non-, light, medium or heavy user of alcohol	self perceived use by sex and form
20	Jan-Jul 1974	Hobart women, 18-60 years	Carrington-Smith (1978)	500	questionnaire administered by trained interviewer	frequency and quantity	frequency only
21	Aug 1974	Ballarat, Vic., (a) secondary students (b) apprentices (c) tertiary students	Graves (1977)	a:739 b:63 c:292	self administered questionnaire	usual drinking pattern, quantity and type	for each group: frequency; number of drinks in one day; combination of these into a QF scale
22	1974	Adelaide secondary schools, 3rd and 5th year students	Quinn *et al* (1975)	455	self administered questionnaire	usual drinking habits: frequency and quantity	frequency by age and sex; quantity (in glasses) on weekdays by age, sex and type of alcohol; similarly for weekends
23	Jun-Aug 1974	Bradbury, a south western Sydney suburb, 18+ years	Strombom (1975)	575	interview questionnaire, usually to housewife	usual frequency and quantity	frequency by quantity; frequency by age and sex; quantity by sex;
24	Oct 1974	Adolescents, 12-17 years, in 30 schools throughout NSW	Egger *et al* (1976)	2741	self administered questionnaire	frequency; type; quantity for each type	frequency and quantity separately, by age and sex
25	1974	Qld schoolchildren, 11-17 years, from 132 schools	Turner and McLure (1975)	3362	self administered questionnaire	usual frequency; quantity on drinking days for beer/wine/spirits/liquer	frequency by grade; quantity by grade

Table 5.6

Australian surveys containing information on the distribution of alcohol consumption
(continued)

	Date	Population	Reference	Sample size	Survey method	Method of estimation of alcohol consumption	Relevant data available
26	Dec 1974 - Aug 1975	Persons admitted to casualty at the Alfred Hospital, Melbourne, following a road crash	Ryan and Salter (1977)	225	interview	usual frequency and quantity	consumption in g/day by BAC [see Table 5.13]
27	1974-75	Elizabeth, SA	Selge (pers. comm.)	1525	questionnaire administered by registered public health nurse	usual drinking habits: frequency and quantity	usual quantity on drinking occasion (grammes) by sex
28	1975	Busselton, WA	Cullen *et al* (1980)	3352	self administered questionnaire to entire population	usual consumption per week (bottles of beer or wine, glasses of spirits)	consumption (g/day) for beer, wine, spirits, by age and sex [see Table 5.14]
29	Apr 1975 - Aug 1976	Adult members of AWU, Sydney	Gibson *et al* (1977)	9829	questionnaire	usual frequency and quantity	frequency by quantity and sex [see Table 5.15]
30	Jun-Dec 1975	All adults undergoing Medicheck screening in Sydney	Reynolds *et al* (1976)	8516	self administered questionnaire on visual display terminal	usual frequency and quantity	frequency by quantity by sex [see Table 5.16]
31	Jan-Nov 1976	All adults undergoing Medicheck screening in Sydney	Reynolds *et al* (1977)	14516	self administered questionnaire on visual display terminal	usual frequency and quantity	frequency by quantity by sex [see Table 5.16]
32	Jan 1976	Young people on beaches near Geelong, plus local youth groups	Barwon Regional Association for Alcohol and Drug Dependence (BRAAD) (1977)	1344	groups on beaches were approached, and if agreeable, completed a self administered questionnaire	7 day recall (last Sunday to previous Monday) plus provision for a "typical" week	number of 7oz beers/day by sex and: age, area of residence, at school or not, living at home or not [see Table 5.17]
33	Jan-Jun 1976	North West region of Melbourne	O'Connell *et al* (1979)	2096	questionnaire administered by 5th and final year medical students	usual frequency and quantity	consumption (g/day) by sex and: age, marital status, country of birth, education, qualification, employment status, occupational status, income [see Table 5.18]
34	Aug 1976	Secondary schools in Geelong area	BRAADD (1977)	6005	self administered questionnaire	7 day recall ("last week") plus provision for "typical" week	number of 7oz beers by sex and age [see Table 5.19]
35	1976	Australian-born students, 17-25 years, at 3 Sydney universities	Sargent (1979)	725	self administered questionnaire	not stated	consumption in qualitative groups by sex
36	late 1976	Employees of 9 companies in Geelong	Graves and Travers (1977); BRAADD (1977); Krupinski (1978)	651	self administered questionnaire	number of drinks on a usual drinking day; number of drinks in a 4 day period: yesterday + last Friday to Sunday; usual intake on each day in a typical week. Converted to qualitative classes using QF index	QF classes by sex; number of drinks on a usual drinking day, in usual week and last week

Table 5.6

Australian surveys containing information on the distribution of alcohol consumption (continued)

	Date	Population	Reference	Sample size	Survey method	Method of estimation of alcohol consumption	Relevant data available
37	1976	Alcohol abusers presenting to the Community Addiction Service, Newcastle, NSW	O'Neill (1977, 1979)	100	interview using the technique 'grogcount'	usual intake	consumption in g/day [see Table 5.20]
38	Feb 1977	Australian population 18+ years	Australian Bureau of Statistics (1978); Milwood and McKay (1978)	15947	interview questionnaire administered as part of ABS Labor Force Survey	7 day recall; for each day, number, type and size of drink was recorded and converted to grammes of alcohol	consumption (g/day) by sex and: age, state, occupation, state capital/other area, marital status, number of cigarettes/day [see Table 5.21]
39	Oct-Nov 1977	NSW adolescents, 12-17 years, at 30 schools	Egger *et al* (1978)	2298	self administered questionnaire	frequency, type, quantity for each type	frequency by age and sex
40	Dec 1977 - Mar 1978	Attendees at Sydney Hospital Health Information and Screening Service	Cooke *et al* (1982)	20920	self administered questionnaire, with assistance of trained sisted	not stated	consumption (units/week) by sex [see Table 5.22]
41	1978	First year Adelaide University students	Baghurst and McMichael (1978)	221	self administered questionnaire	7 day recall plus usual intake	consumption (g/day) by sex [see Table 5.23]
42	1978	Busselton, WA	Cullen *et al* (1980)	4002	self administered questionnaire to entire population	frequency, plus details of an average week's consumption	consumption (g/day) for beer, wine, spirits, by age and sex [see Table 5.24]
43	1978-79	Non-officer RAAF recruits, Edinburgh SA, (a) at beginning, (b) at end of course	Baghurst and McMichael (1978); Baghurst (pers. comm.)	a:563 b:611	self administered questionnaire	7 day recall plus usual weekly intake	consumption (g/day) by age [see Table 5.25]
44	1978-79	Perth male "social drinkers"	Tofler and Woodings (1981)	320	interview	not stated	consumption (l. beer or equivalent/day) [see Table 5.8]
45	July 1979	Tasmanian high school students (ages 12-16)	Lynch *et al* (1981)	1211	questionnaire	7 day recall	categorised consumption of drinker (never, past, present light, present heavier) by age and sex
46	Jan - June 1980	Qld human-service students	Engs (1982)	1449	questionnaire	average frequency plus usual amount consumed per occasion in past year	consumption (g/day) by sex; by year of study; by course; by perceived importance of religion [see Table 5.26]
47	June 1980	residents of Townsville, aged 15-85	Grichting (1983)	303	interview	recall of last 24 hours + last weekend. If atypical, usual consumption used.	consumption (g/day) by age and sex [see Table 5.27]

Table 5.7

Alcohol consumption - Heyfield, 1965

Consumption (g/day)	Age 14-21	22-64	65+	Total
Males				
0	85	50	12	147
≤5.4	25	118	15	158
5.4-32.6	7	154	12	173
>32.6	5	130	1	136
Total	122	452	40	614
Females				
0	81	115	20	256
≤5.4	24	166	7	197
5.4-32.6	-	68	3	71
>32.6	-	7	7	14
Total	105	396	30	538

Notes:

References: Krupinski *et al* (1967); Krupinski *et al* (1970).

Consumption unknown: 12 males, 7 females.

The class limits were originally 1-5, 5-30, 31+ glasses of beer/week, and were converted by Krupinski (1978) to 1-7.15, 7.15-42.9, 42.9+ g/day. The error in this conversion, based on 10 g alcohol/200 ml glass beer, was pointed out by the Australian Associated Brewers (1978). The present limits were calculated from the data in Table 5.2.

Table 5.8

Alcohol consumption – Perth male "social drinkers"

Consumption (g/day)	1965–66	1978–79
0	72	67
1–36	95	91
37–72	72	74
73–135	54	46
>135	66	42
Total	359	320

Notes

References: Saker *et al* (1967), Tofler *et al* (1969), Tofler and Woodings (1981).

Subjects for the survey were obtained by "visiting a number of hotel managers in the Perth metropolitan area, and asking them to volunteer for a medical examination. In turn these men asked their friends, colleagues and business associates to take part. To be accepted for the survey, subjects had to be fully employed males who considered themselves healthy, particularly with regard to the cardiovascular system." (Tofler and Woodings, 1981).

Alcohol intake was converted to equivalent volume of beer intake per day. Conversion factors used were 3 for light wine, 5 for fortified wine, and 10 for spirits. Conversion to g/day is on the basis that W.A. beer contains approximately 4% alcohol by volume. (Saker *et al*, 1967).

Table 5.9

Alcohol consumption – Alcoholism Clinic patients, 1966–67

Consumption (g/day)	Males	Females
<100	0	0
100–150	39	18
150–200	20	2
200–250	43	4
250–300	15	3
300–350	9	1
350–400	6	–
400–450	6	–
450–500	2	–
>500	3	–
Total	143	28

Notes

Reference: Wilkinson *et al* (1969).

Data is taken from a photographic enlargement of Figure 7 of the reference. The authors state that there were 179 men and 41 women but "data obtained from 38 men and 12 women were incomplete or unreliable in part". This leaves 141 men and 29 women, which agrees fairly closely with figures as taken from the histogram.

Table 5.10

Alcohol consumption – Prahran, 1968

Consumption (g/day)	Age 0-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	Total adults (20+)	Total all
Males										
0	96	74	35	16	15	14	21	15	116	286
1-10	7	33	68	40	25	28	21	18	200	240
11-40	–	8	83	49	44	32	25	16	249	257
41-80	–	0	33	18	10	12	6	5	84	84
81-120 120+	–	1	17	12	11	4	14	1	34 25	60
Total	103	116	236	135	105	90	87	55	708	927
Females										
0	116	103	78	34	37	26	39	61	275	494
0-10	14	55	160	55	41	49	38	42	385	454
11-40	–	8	54	23	25	35	16	13	166	174
41-80	–	–	1	1	3	4	1	–	10	10
81-120 120+	–	–	2	5	1	–	1	–	6 3	6 3
Total	130	166	295	118	107	114	95	116	845	1141

Notes:

Reference: Rankin and Wilkinson (1971).

Consumption unknown: 34 men, 22 women (adults)

The consumption was calculated on the basis of 10g alcohol being contained in "one 7 oz glass of beer, 1 fluid ounce of whisky or gin, or two fluid ounces of sherry". This conversion is incorrect (see Table 5.2), but as the mix of drink type is unknown, it is impossible to correct it.

Table 5.11

Alcohol consumption – Bourke aborigines, 1972

Consumption (g/day)	age 10-19	20-29	30-39	40-49	50+	Total adults(20+)	all
			Males				
0	64	0	2	4	6	12	76
1-10	10	5	8	0	1	14	24
11-40	4	6	2	1	3	12	16
41-80	2	8	5	1	3	17	19
81-120	–	8	8	4	2	22	22
121-180	–	19	9	6	4	31	31
180+	–	2	3	3	5	13	13
Total	80	41	37	19	24	121	201
			Females				
0	80	36	26	14	15	91	171
1-10	–	3	3	3	3	12	12
11-40	–	7	0	5	1	13	13
41-80	–	3	2	0	0	5	5
81-120	–	–	0	1	0	1	1
121+	–	–	1	–	2	3	3
Total	80	49	32	23	21	125	205

Notes:

Reference: Kamien (1975b).

Consumption unknown: 3 males, 3 females.

Table 5.12

Alcohol consumption - South East S.A., 1972-73

Consumption (g/day)	Persons
0	95
1-10	388
10-30	121
30-80	10
80+	4
Total	618

Notes:

Reference: Selge (1975).

The data have been converted from a quantity-frequency state to an absolute alcohol state as follows: We assume

rarely	= 0.1-0.9/month
once/month	= 1/month
once/fortnight	= 2-3/month
once/week	= 4-8/month
every few days	= 9-27/month
daily	= 28-30/month

Selge gives the quantity as grammes of alcohol. The various frequency-quantity states given in Selge's Table 10:9 are then converted to g/day as shown in Table 5.12a (Intakes above 300g were not included in Table 10:9, and were kindly identified by Dr. B. Selge).

Table 5.12a

Conversion of QF data to g/day - South East S.A.

Consumption group	Frequency	Quantity (g)	Daily consumption (g) min	mean	max	No. of persons
0	nondrinkers		0.0	0.0	0.0	95
< 10	rarely	1-10	.0	.1	.3	102
	1/mnth	1-10	.0	.2	.3	7
	1/fnt	1-10	.1	.4	.1	8
	rarely	10-50	.0	.5	1.5	103
	1/mnth	10-50	.3	1.0	1.7	33
	1/week	1-10	.1	1.0	2.7	7
	rarely	50-100	.2	1.3	3.0	9
	1/fnt	10-50	.6	2.5	5.0	22
	1/mnth	50-100	1.7	2.5	3.3	2
	rarely	100-200	.3	2.5	6.0	0
	ev.few days	1-10	.3	1.3	3.0	9
	rarely	200-300	.7	4.2	9.0	2
	daily	1-10	.9	4.8	10.0	11
	1/mnth	100-200	3.3	5.0	6.7	0
	1/week	10-50	1.3	6.0	13.3	66
	1/fnt	50-100	3.3	6.3	10.0	2
	rarely	500+	1.7	8.3+	-	1
						388
10-30	1/fnt	100-200	6.7	12.5	20.0	0
	1/week	50-100	6.7	15.0	26.7	11
	ev.few days	10-50	3.0	18.0	45.0	54
	daily	10-50	9.3	29.0	50.0	56
	1/week	100-200	13.3	30.0	53.3	0
						121
30-80	ev.few days	50-100	15.0	45.0	90.0	7
	daily	50-100	46.7	72.5	100.0	3
						10
> 80	ev.few days	100-200	30.0	90.0	180.0	1
	1/week	400-500	53.3	90.0	133.3	1
	daily	100-200	93.3	145.0	200.0	2
						4

Table 5.13

Alcohol consumption – road crash victims, 1974–75

Consumption (g/day)	**BAC (g%)** 0–0.049	≥0.05	Total
0	13	1	14
<1	33	0	33
1–10	53	14	67
11–40	38	33	71
41–80	6	18	24
81–120	2	9	11
>120	2	3	5
Total	147	78	225

Notes:

Reference: Ryan and Salter (1977).

Table 5.14

Alcohol consumption – Busselton, W.A., 1975

Consumption (g/day)	Age < 30	30–39	40–49	50–59	60–69	70+	Total
Males							
0	82	56	70	62	115	73	458
1–20	149	114	102	104	82	66	617
21–40	51	56	58	56	62	32	315
41–60	23	19	24	27	22	10	125
61–80	5	8	1	4	5	5	38
81–100	5	1	1	5	4	5	21
100+	1	5	4	3	3	3	19
Total	316	259	270	261	293	194	1593
Females							
0	193	160	156	184	192	134	1019
1–20	122	99	111	101	66	39	538
21–40	22	22	24	34	25	8	135
41–60	8	4	8	9	12	3	44
61–80	1	2	2	4	3	0	12
81–100	0	0	0	–	2	1	3
100+	1	2	2	–	1	2	8
Total	347	289	303	332	301	187	1759

Notes:

Reference: Cullen *et al* (1980).

Alcohol intake calculated on the basis that one 750 ml bottle of beer contains 29.5 g alcohol, one 750 ml bottle of wine contains 90 g alcohol and one bottle of spirits averages 340 g alcohol per litre (sic).

Table 5.15

Alcohol consumption – AWU members, 1975–76

Consumption (g/day)	Males	Females
0	401	1691
<10	1256	3876
10–40	1164	1157
40–80	172	27
80+	70	14
Total	3063	6765

Notes:

Reference: Gibson *et al* (1977).

Data has been converted from QF scale as follows:
Quantity:
Since the most common size of a Sydney glass of beer is 285 ml, 1 glass has been assumed to contain 10 g alcohol.
Frequency:

very rarely	= 1/8 per week
once a week	= 1 per week
couple of times per week	= 2–4 per week
most days	= 5–7 per week.

Table 5.16a shows the conversion.

Data from both Medicheck samples (Reynolds *et al*, 1976, 1977) has been converted using Table 6.16a also.

Table 5.16

Alcohol consumption - Medicheck screenings, 1975, 1976

Consumption (g/day)	1975 Males	1975 Females	1976 Males	1976 Females
0	1140	1719	1739	3112
1-10	1001	861	1801	1821
10-40	2100	897	3270	1546
40-80	376	47	639	66
80+	176	11	246	13
Total	4793	3535	7695	6558

Notes

References: Reynolds *et al* (1976, 1977).

The data have been converted from QF classes to absolute alcohol as in Table 5.16a.

Table 5.16a

Conversion of QF data to g/day - AWU workers and Medicheck

Consumption group	Frequency	Quantity (drinks)	Daily consumption (g) min	mean	max	AWU %	Males no.
0	nondrinkers		0.0	0.0	0.0	13.1	401
< 10	very rarely	1-2	.2	.3	.4	12.1	
	very rarely	3-5	.5	.7	.9	2.6	
	very rarely	6-8	1.1	1.2	1.4	0.3	
	very rarely	> 9	1.6	1.6+	-	0.1	
	once/week	1-2	1.4	2.1	2.8	8.3	
	once/week	3-5	4.3	5.7	7.1	6.3	
	couple/week	1-2	2.8	86.4	11.4	10.3	
	once/week	6-8	8.6	10.0	11.4	1.1	
							1256
10-40	most days	1-2	7.1	12.9	20.0	7.9	
	once/week	> 9	12.9	12.9+	-	0.7	
	couple/week	3-5	8.6	17.1	28.6	10.6	
	couple/week	> 9	25.7	25.7+	-	1.0	
	couple/week	6-8	17.1	30.0	45.7	3.1	
	most days	3-5	21.4	34.3	50.0	14.7	
							1164
40-80	most days	6-8	42.9	60.0	80.0	5.6	172
80+	most days	> 9	77.1	77.1+	-	2.3	70

Table 5.17

Alcohol consumption - Geelong beach survey, 1976

Consumption (g/day)	Age 10-14	15-19	20-24	25+	Total
Males					
0	25	134	31	2	192
1-30	24	212	49	32	317
31-60	4	69	16	10	99
61-90	-	49	29	7	85
90+	-	25	11	6	42
Total	53	491	136	57	735
Females					
0	48	148	18	9	223
1-30	25	232	55	28	340
31-60	3	16	17	2	38
61-90	-	4	2	-	6
90+	-	4	-	-	4
Total	76	400	92	39	607

Notes:

Reference: Barwon Regional Association for Alcohol and Drug Dependence (1977).

Figures in the body of the table are derived from published percentages and so many may not agree with totals. Totals for age groups are correct. Consumption was given in "7 oz glasses beer/day" and has been converted to grammes/day on the basis of 7 oz beer = 7.6 g alcohol.

Table 5.18

Alcohol consumption – North West Melbourne, 1976

Consumption (g/day)	Age 15-19	20-29	30-39	40-49	50-59	60-69	70+	Total
				Males				
0	54	23	21	20	15	17	15	165
1-9	65	124	88	71	58	40	20	466
10-39	17	44	45	29	33	17	7	192
40-79	7	30	29	34	19	15	4	138
80+	2	10	10	21	7	6	1	57
Total	145	231	193	175	132	95	47	1018
				Females				
0	60	49	49	47	35	43	27	310
1-9	80	176	123	107	89	48	34	657
10-39	5	17	24	12	14	11	5	88
40-79	–	6	2	10	3	1	–	22
80+	–	1	–	–	–	–	–	1
Total	145	249	198	176	141	103	66	1078

Notes:

Reference: O'Connell *et al* (1979).

Details of conversion from a QF measure to g/day are given by Krupinski (1978). His assumptions are:

1. a standard drinks contains 10 g alcohol
2. most days = 20 days/month
3. weekends = 8 days/month
4. social occasions = 1 day every 2 months
5. rare occations = 1 day every 2 months
6. mean daily consumption = mean no. of drinks per occasion × no. of occasions per month ÷ 30.

Table 5.19

Alcohol consumption - Geelong School Survey, 1976

Consumption (g/day)	**Age** 14	15	16	17	18	"Other"	Total
				Males			
0	565	492	365	126	26	36	1609
1-8	209	295	183	83	15	10	795
9-15	64	86	95	56	18	4	322
16-30	17	80	60	27	15	1	201
31-60	13	19	26	10	3	1	72
61-90	3	13	3	5	1	-	25
91+	1	3	4	1	-	-	9
Total	872	987	736	308	78	52	3033
				Females			
0	640	598	397	164	20	26	1845
1-8	189	291	204	118	20	9	831
9-15	28	67	65	39	4	3	206
16-30	11	20	20	17	-	1	69
31-60	4	4	1	4	-	1	14
61-90	0	4	2	-	-	-	6
91+	1	-	-	-	-	-	1
Total	873	984	689	342	44	40	2972

Notes:

Reference: Barwon Regional Association for Alcohol and Drug Dependence (1977).

"Other" includes ages 11-13, 19 or not stated.

Table 5.20

Alcohol consumption - Newcastle alcohol abusers, 1976

Consumption (g/day)	No. of persons
< 80	0
80-120	1
120-180	47
180-240	19
240-300	15
300-360	11
360-420	1
420-540	0
540-600	1
600-660	2
660-720	0
720-780	2
780-800	1
Total	100

Notes:

Reference: O'Neill (1977).

Table 5.21

Alcohol consumption – ABS survey, 1977

Consumption (g/day)	**Age** 18–24	25–44	45–64	65+	Total
Males					
0	355	645	613	345	1959
1–9	328	750	527	223	1829
10–19	204	523	325	102	1153
20–29	167	450	275	54	946
30–39	103	239	173	31	545
40–49	100	189	134	18	441
50–59	34	160	104	16	313
60–69	39	85	77	7	208
70–79	28	66	35	7	136
80–89	↑	34	26	↑	71
90–99	|	34	17	|	67
100–109	|	17	↑	|	38
110–119	61	22	|	6	42
120–149	|	28	60	|	52
150–199	|	14	|	|	37
200+	↓	7	↓	↓	19
Total	1419	3263	2366	809	7856
Females					
0	651	1402	1275	796	4125
1–9	519	1130	672	260	2580
10–19	139	400	223	46	808
20–29	55	119	100	23	297
30–39	27	51	61	9	148
40–49	7	33	16	0	56
50–59	1	12	15	2	30
60+	9	20	13	5	47
Total	1408	3167	2375	1141	8091

Notes:

Reference: Australian Bureau of Statistics (1978).

The ABS publication does not give sample values for the distributions, but expresses their results in terms of the total Australian population. The total sample size was 15947, split amoung the states as follows: NSW 3885; Vic. 3395; Qld. 2435; SA 2232; WA 2188; Tas. 1157; NT 145;

ACT 510 (D. Seal, ABS, pers. comm.). The values in the table above have been calculated on the basis of the sampling fractions for age and sex being equal to the population values, which is a reasonable assumption according to Seal. The values in the table are in reasonable agreement with those calculated from the sum of the individual state categories, particularly at higher consumption levels:

Consumption	Total persons calculated from age-sex breakdown	Total persons calculated from state values
0	6084	5900
1-39	8306	8482
40-79	1207	1210
80+	350	353
Total	15947	15947

Table 5.22

Alcohol consumption - December 1977 to March 1981
Sydney Hospital Health Information and Screening Service

Consumption (g/day)	Males	Females
0	3061	3423
1-9	3004	2109
10-29	4128	1526
>30	3342	327
Total	13535	7385

Notes

Reference: Cooke *et al* (1982).

Cooke *et al* give consumption in units per week, where "one unit = 10 g of ethanol = one glass of beer".

Table 5.23

Alcohol consumption - Adelaide University students, 1978

Consumption (g/day)	Males	Females
0	51	16
1-10	74	39
11-20	29	5
21-30	3	2
31-40	1	-
41+	1	-
Total	159	62

Notes:

Reference: Baghurst (pers. comm.)

Table 5.24

Alcohol consumption – Busselton W.A., 1978

Consumption (g/day)	Age <30	30–39	40–49	50–59	60–69	70+	Total
Males							
0	92	67	69	99	126	115	568
1–20	114	127	111	97	110	94	653
21–40	83	66	54	45	40	43	331
41–60	30	28	19	39	27	12	155
61–80	7	10	5	15	14	4	55
81–100	7	2	10	8	6	2	35
100+	8	4	6	5	5	3	31
Total	341	304	274	308	328	273	1828
Females							
0	204	159	167	234	217	192	1173
0–20	174	177	131	129	125	68	804
21–40	21	17	27	42	25	10	142
41–60	3	8	7	10	13	1	42
61–80	3	0	3	–	2	1	9
81–100	–	1	–	–	1	–	2
100+	–	–	–	–	2	–	2
Total	405	362	335	415	385	272	2174

Notes:

Reference: Cullen *et al* (1980).

Alcohol conversion on the basis of one 750 ml bottle of beer = 29.5 g alcohol, one 750 ml bottle wine contains 90 g alcohol, and alcohol content of spirits is either 340 g/l or 260 g/l.

Note: the paper by Cullen *et al* contains a misprint in Table 5. The frequency of total beverage consumption for agegroup 50–59, consumption group 100+ g/day, should be 5, not 6. This error has been corrected here.

Table 5.25

Alcohol consumption - RAAF recruits, 1978-79

Incoming recruits

Consumption (g/day)	Age 17-20	21-25	26+	Unknown	Total
0	86	15	16	2	119
<30	203	66	52	5	326
30-50	33	23	7	1	64
50-70	16	11	5	1	33
70-90	4	2	3	0	9
>90	4	5	3	0	12
Total	346	122	86	9	563

Outgoing recruits

Consumption (g/day)	Age 17-20	21-25	26+	Unknown	Total
0	88	9	18	5	120
<30	182	73	42	7	304
30-50	60	26	13	3	102
50-70	21	14	10	0	45
70-90	10	3	3	0	16
>90	13	5	5	1	24
Total	374	130	91	16	611

Notes:

Reference: Baghurst and Dwyer (1981); numerical data: Baghurst (pers. comm.)

Table 5.26

Alcohol consumption – Queensland human-service students, 1980

Consumption (g/day)	Males	Females
0	118	63
1-19	469	512
20-39	127	55
40-59	42	12
60-79	23	2
≥80	24	2
Total	803	646

Notes

Reference: Engs (1982).

Alcohol conversion on the basis of each 10 oz (285 ml) beer was considered to contain 10.4 g alcohol, each wine glass of wine (90 ml) was considered to contain 8.2 g, and each "nip" (30 ml) of distilled spirits to contain 9.2 g of absolute alcohol.

Factors used in calculating the amount of beverage consumed were:

every day	365
3 or 4 times a week	182
1 or 2 times a week	78
2 to 4 times a month	34
2 or 3 times a year	3.5
about once a year	1
used or experimented with	0.1
never used	0

Table 5.27

Alcohol consumption – Townsville residents, 1980

Consumption (g/day)	Age 15-17	18-24	25-44	45-64	65+	Total
Males						
0	6	7	13	9	6	41
1-40	5	13	20	12	6	56
41-80	1	17	11	4	5	38
81-120	0	3	10	2	1	16
>120	0	2	3	1	0	6
Total	12	42	57	28	18	157
Females						
0	10	12	31	18	8	79
1-40	2	21	24	14	8	69
41-80	0	1	2	4	1	8
81-120	0	0	0	0	0	0
>120	0	1	0	0	0	1
Total	12	35	57	36	17	157

Notes

Reference: Grichting (1983), Grichting (pers. comm.).

No information available for 44 of the total sample size of 358 respondents.

Alcohol conversion on the basis of

light beer	2.4%
regular beer	4.8%
table wine	15.0%
fortified wine	18.0%
spirits	40.0%

Chapter 6

Results.

6.1 Scope of analyses

The previous chapter listed data from 21 Australian surveys on alcohol consumption. In this study, the main interest lies in the examination of data from samples of what could loosely be termed "typical Australians". Three surveys (Alcoholism Clinic patients, Bourke aborigines, and Newcastle alcohol abusers) were considered to represent atypical populations and have been excluded from the comments and analyses of this chapter.

It is possible to envisage analyses of greater complexity than are presented in this chapter, but this has not been done for several reasons:

1. In general, the quality of survey data on alcohol consumption does not warrant complex analysis. The previous chapter discussed the large discrepancies between reported and apparent consumption of alcohol. There is little point in building complex constructions on poor foundations.

2. It has been shown consistently that the distribution of alcohol consumption is "unimodal, continuous, positively skewed, and similar to a log-normal distribution" (see Chapter 2) and there are no a *priori* reasons why Australian data should be grossly different to that overseas. The analyses to be presented are sufficient, so far as the data will allow it, to detect discrepancies from this description.

As is usual in alcohol studies, we will be concerned primarily with data about consumers of alcohol and ignore data about abstainers. We therefore take a brief look at abstainers now.

6.2 Abstainers

Table 6.1 shows the percentage of abstainers from adult samples in the surveys. The data are shown graphically in Figure 6.1. In recent years there is no strong discernible trend in the proportion of abstainers, the most notable feature of the figures being the variation. This reflects the diverse nature of the populations sampled.

Table 6.1

Percentage of adult abstainers.

Survey	date	percent abstainers male	female	ratio
Heyfield	1965	12.6	31.7	0.40
Perth social drinkers	1965-66	20.1	-	-
Prahran	1968	16.4	32.5	0.50
South east SA	1972-73	15.4		-
Roadcrash victims	1974-75	6.2		-
Busselton	1975	28.8	57.9	0.50
AWU members	1975-76	13.1	25.0	0.52
Medicheck	1975	23.8	48.6	0.49
Medicheck	1976	22.6	47.5	0.48
Geelong beach	1976	17.4	20.6	0.84
NW Melbourne	1976	12.7	26.8	0.47
ABS survey	1977	24.9	50.9	0.49
Sydney Hospital	1977-81	22.6	46.4	0.49
Adelaide Univ.	1978	32.1	25.8	1.24
Busselton	1978	31.1	54.0	0.58
RAAF ingoing	1978-79	21.1	-	-
RAAF outgoing	1978-79	19.6	-	-
Perth social drinkers	1978-79	20.9	-	
Qld students	1980	14.7	9.8	1.50
Townsville	1980	24.1	47.6	0.51

However with three exceptions, the ratio of male to female abstainers shows a remarkable stability, at approximately 0.5. Two of these three samples are student populations, and are notable for the fact that they are the only samples in which there is a greater proportion of female drinkers than male

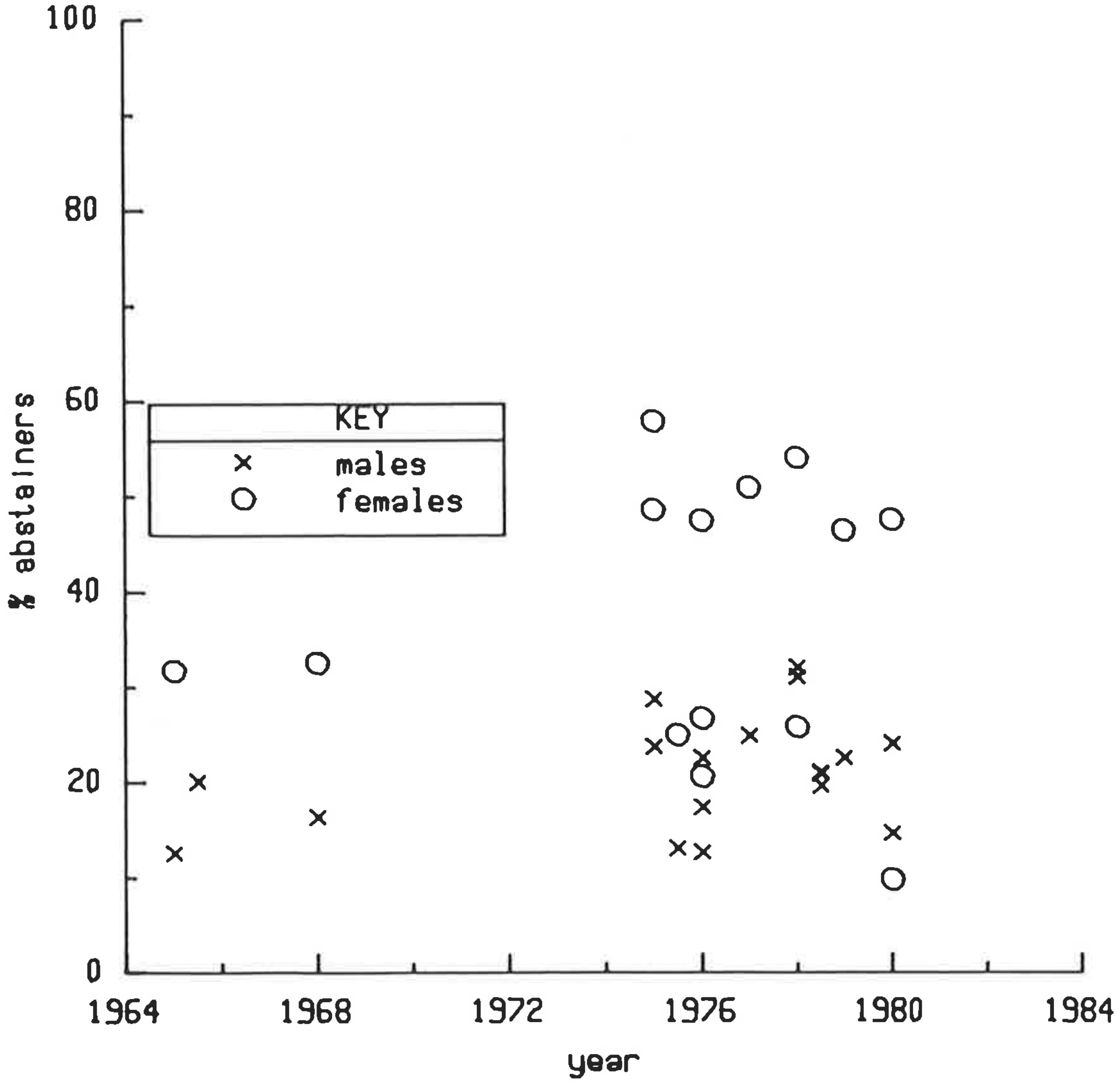

Figure 6.1 Percentage of adult abstainers.

drinkers.

6.3 Consumers – sample statistics

The remainder of this chapter will be concerned with data conditional on a non-zero consumption. Table A1 lists sample statistics from each of the 21 surveys. Where possible, details are given separately for each agegroup and sex. Additionally, if the survey included youths, statistics are given for both the adult and youth portions of the sample. The definition of "youths" varies for each sample, depending on which agegroups the experimenter has used in presenting the results. The actual agegroups used are noted in Table A1.

The statistics listed in the table are the sample size, the number of classes into which the total sample is grouped, the mean (in g/day) of the individual alcohol consumption, the standard deviation of the logarithm of the alcohol consumption, and the skewness of the sample. Since the class interval widths differ from survey to survey, Shepherd's correction has been applied to the standard deviation and skewness, to remove the grouping bias (Bliss, 1967).

In most cases the upper class interval has only a lower bound, say x_m. The usual assumption has been made, namely the midpoint of the uppermost class interval is taken to be the same distance above x_m as the midpoint of the previous class interval is below it.

6.3.1 <u>Sample sizes and groupings</u> Sample sizes (including abstainers) range from about two hundred to in excess of twenty thousand, and reflect the methodology used in carrying out the survey (see Table 5.6). Those surveys with small sample sizes usually used an interview to obtain information, while the larger ones tended to use self-administered questionnaires, sometimes with input directly to a computer via a visual display terminal.

Once abstainers are removed from the sample, the numbers of consumers in the samples range from 154 to 14709, with a median of 1267. Corresponding figures for male consumers in the samples are 108, 10747 and 684, while for female consumers, the minimum, maximum and median are respectively 46, 5074 and 768. Thus the sample sizes are positively skewed; that is, there are more samples with sizes towards the lower end of the range.

In considering the distribution of consumption, there are good reasons to think that the consumption patterns of different agegroups may differ. When we consider that the sample sizes referred to in the previous paragraph may subsequently be divided among four to six agegroups, we see that in many cases, sample sizes in age x sex subgroups are very small. Table A1 shows that they range from 2 to 2618. The median size is 125.

Data on alcohol consumption is usually published as grouped frequency data, typically with four to six classes. Common choices for class intervals (in g alcohol/day, rounded to the nearest gramme) are

1-10, 11-40, 41-80, 81+

1-10, 11-40, 41-80, 81-120, 121+

1-20, 21-40, 41-60, 61-80, 81-100, 101+

Additionally, there are often no observations in the upper class intervals for some age x sex subgroups. Of the 18 surveys under consideration, there were 2,7,4 and 4 surveys with 3,4,5 and 6 classes respectively, and one (ABS, 1978) where consumption was subdivided into 15 classes, although in

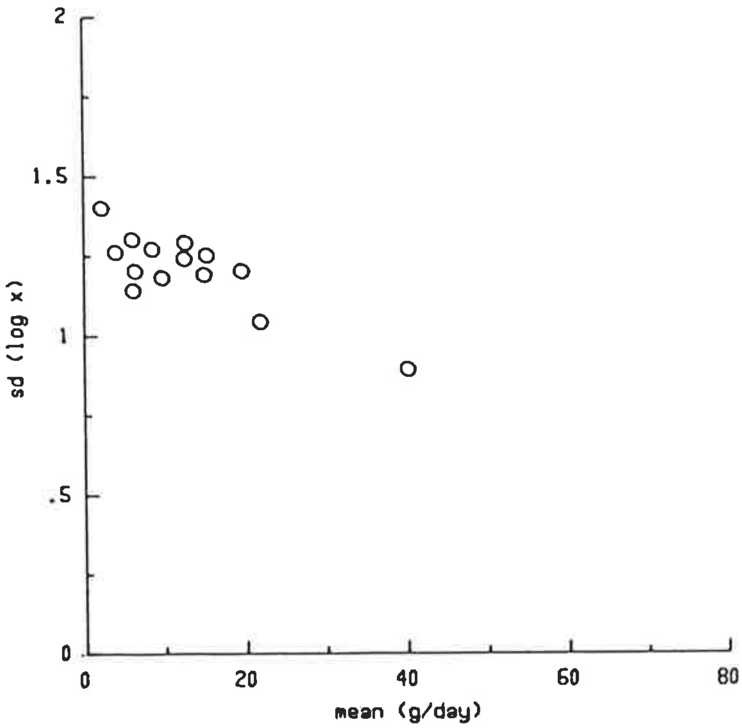

Figure 6.2 Figure 2 of Bruun *et al* (1975).

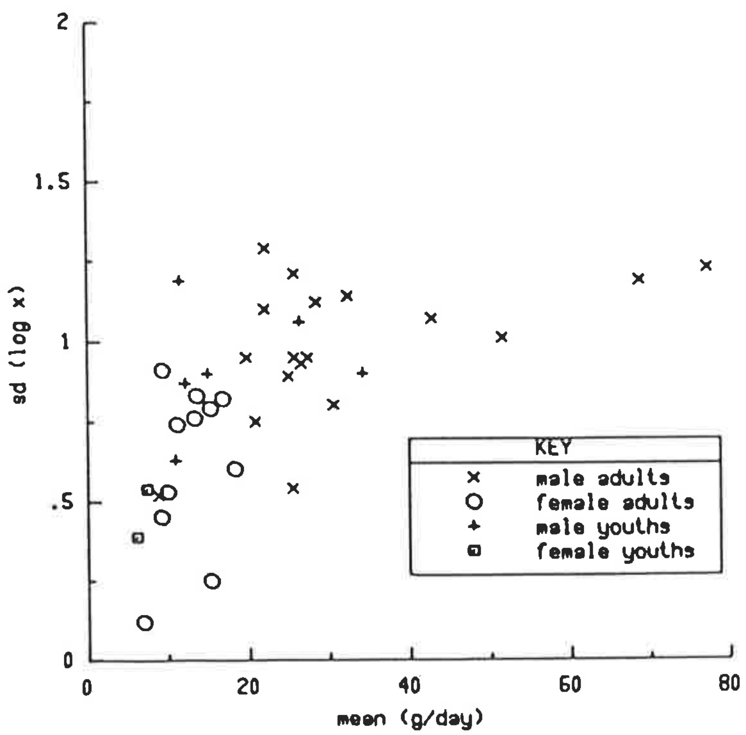

Figure 6.3 Standard deviation of samples.

this latter survey, there are only 7 classes for each of the female agegroups, and only one male agegroup has data for the full 15 classes.

6.3.2 Mean consumption and standard deviation As may be expected from such a diverse group of surveys, the values for mean consumption cover a wide range. The agegroup means for males range from 11.1 to 56.4 g/day; the corresponding range for females is 2.7 to 28.8 g/day.

Since the age subdivisions used in the surveys are not standardised, it is difficult to be specific about any overall trends with age. There is a suggestion in some samples (ABS, NW Melbourne and others) of a quadratic response with age. That is, lower consumption in the lower agegroups, increasing in the central agegroups, and finally decreasing again in later agegroups.

Taking adult samples only, the range for the means of the male samples is 8.9 to 77.3, and for females 7.0 to 25.8 g/day.

Bruun *et al* (1975) examined the relation between the mean consumption and the standard deviation of the logarithm of consumption using data from European surveys. The figure they used to do this has been the subject of much discussion in the literature, and has been reproduced several times (Duffy, 1977a; Duffy and Cohen, 1978; Skog, 1980a, 1983; Mohan *et al*, 1980). It is reproduced again here as Figure 6.2. (Bruun *et al* gave the consumption in terms of litres of alcohol per year; it has been converted to grammes per day in Figure 6.2). The 14 data points represent 6 adult populations, plus 8 subgroups from 2 surveys of Scandinavian youths. It is of interest to compare it with a similar figure, Figure 6.3, derived from Australian data. The data for Figure 6.3 is taken from Table A1, and comprises the samples for both adults and youths of both sexes. These four categories are differentiated on the figure.

It is evident from Figure 6.3 that the samples of females have, in general, both lower mean consumption and lower standard deviation than the samples of males. The other striking feature is that the scatter of points in Figures 6.2 and 6.3 is very different. While the data of Bruun *et al* demonstrated a negative correlation between the two statistics, the Australian data is positively correlated. Bruun *et al* stated, by visual inspection of their figure, that "differences as to dispersion between populations with similar levels of consumption are quite small". This conclusion is obviously quite inappropriate for the Australian data.

6.3.3 Skewness All samples exhibit positive skewness (see Table A1), indicative of the preponderance of light and medium drinkers in the samples. (Most skewness coefficients are significantly greater than zero; however in the present instance there is little interest in testing departure from normality). Comparisons within the one survey of corresponding agegroups for males and females show that the skewness is greater for females than males, with very few exceptions. The reason for this is that although both male and female consumption is distributed over a similar range, the lower mean consumption for females means that the distribution for females is more "squashed" to the left than that for males.

6.4 Fitted distributions

6.4.1 Introduction In Chapter 7 we show that the maximum likelihood estimation of distributions from grouped data can be formulated as an iterated weighted regression. The method requires specification of the first derivatives of the class probabilities with respect to the parameters. The necessary derivatives for fitting the two and three parameter lognormal distribution, both untruncated and truncated, were given in Section 4.4.

Tables A2 to A24 give information about various lognormal and gamma distributions fitted to 19 of the 21 data sets given in Chapter 5. All lognormal distributions were fitted using the above method. Programs were written using Matlab (Moler, 1976) and run under the Unix operating system on a DEC Vax 11/750 computer. Gamma distributions were fitted using the program MLP (Ross, 1980), on the same computer.

The tables in the Appendix list, for each fit, the number of observations, the parameter estimates, the χ^2 goodness-of-fit statistic and the probability of its significance. The log-likelihood ratio χ^2 statistic (Fisher, 1950b), rather than the more usual Pearson statistic, has been used, since it requires no pooling of the tail frequencies, which in this case are of considerable interest (Bliss, 1967). Larntz (1978) has shown that the test tends to be conservative in comparison with the Pearson statistic. Other information has been omitted in the interests of trying to make a large amount of information more readable. The tables are not exhaustive, in that details of fits of all possible lognormal and gamma distributions are not given for every data set, although the two simplest models, the two parameter lognormal and the gamma distributions, have been fitted to most agegroups for all data sets. For instance, in Table A9 (Busselton 1975, females) details are given for fits of the two parameter lognormal to each agegroup, but only for adults for the same distribution censored at 40 g alcohol/day. This is occasionally because of the difficulty in finding adequate starting values for a particular fit, but usually just because it was felt that including the extra detail serves no useful purpose. In the particular case of Table A9, since the (unrestricted) two parameter lognormal is an adequate representation of the data in each agegroup, it was felt unnecessary to supply details for each agegroup for the censored fit as well. (In fact, censored fits with non-significant χ^2 values exist for all agegroups except ages 50-59, which does not have enough class intervals to permit the censored distribution to be fitted.)

Before examining the fits we note that two data sets (Heyfield, Table 5.7; Sydney Hospital, Table 5.22) contain only three classes of consumers. This is not enough to permit the fitting of a two parameter distribution. An exponential distribution (one parameter) was fitted to both these data sets, but the only fit which gave a non-significant χ^2 value was Heyfield males aged over 65 (parameter value 0.123, χ^2_1 = 0.79). This agegroup contains only 28 observations, with one class interval containing only one of these.

In attempting to summarise the results of nearly 30 pages of tables covering nearly 400 distributional fits, it is neither helpful nor of interest to give a detailed description of each fit. Our primary interest is in the elucidation of overall patterns. As a first stage we will ignore data about individual agegroups, and consider fits to aggregate adult agegroups, where they exist.

6.4.2 Fits to aggregate adult groups Seven of the data sets contain only four class intervals, permitting the fitting of only the two parameter lognormal and the gamma distributions. Of these seven, neither specification gives an adequate fit to the data from South East S.A. (Table A6), to the AWU members, both sexes (Table A10), to the Medicheck data, again both male and female (Table A11), to male data from the Geelong Beach study (Table A12), or to male data from the North-west Melbourne study (Table A13). Both distributions are acceptable as descriptions of the Geelong Beach female data (Table A12), and both male and female data for the Townsville residents (Table A24). The lognormal, but not the gamma distribution fitted both year's data from the Perth social drinkers (Table A2) and the opposite situation held for the North-west Melbourne females (Table A13). In cases where neither specification fitted, neither the exponential nor the Weibull distributions could provide a better fit. Given the small numbers of class intervals in all these data sets, it is impossible to try fitting models such as the three parameter lognormal, or to fit a two parameter distribution to the data with some of the lower classes censored. As we will see later, both of these alternatives have merit if our interest lies predominantly in the upper tail of the distributuion.

The remaining data sets (that is, the ones with five or more class intervals), show no particular predisposition to either a two parameter lognormal or gamma specification. Of the six data sets for females, three (Busselton 1975 and 1978, Tables A9 and A21; Queensland students, Table A23) are adequately described by both the lognormal and gamma distributions, and three (ABS, Table A18; Geelong School study, Table A15; Prahran, Table A4) by neither. Both specifications fit the data for males from the surveys of Adelaide University (Table A19), RAAF trainees (Table A22) and Queensland students (Table A23), while for only one data set (ABS, Table A17) will neither specification suffice. The lognormal but not the gamma describes the male samples from the Geelong School survey (Table A14), Busselton 1975 (Table A8) and Prahran (Table A4); the reverse holds for the Busselton 1978 males (Table A20).

These data sets, however, do allow more freedom with the choice of a specification. One possibility is to use a specification depending on more than two parameters. A three parameter lognormal distribution gave a non-significant value of χ^2 for all adult samples mentioned above, except for ABS males.

A second possibility is to use either a censored or a truncated two parameter lognormal distribution. In fact the grouping of observations amounts to a censoring of the distribution; what we consider here is to amalgamate the first two (or more if the number of class intervals permits) class intervals, and refit the two parameter lognormal with the reduced number of class intervals. In doing this, we are censoring information in the lower tail without discarding it entirely, and might reasonably expect the fit to be more dependent on the shape of the upper tail. On the other hand, truncation of one or more of the lower class intervals, although leaving the fit more dependent on the upper tail, discards much of the available information. Graphically we can demonstrate the difference between the censored and uncensored fits by means of a log-probability plot. Figure 6.4 shows a plot for the ABS adult male data. The solid line represents the uncensored fit to the data (χ^2 = 236, P < 0.001). The points at the lower end of the line, having greater weight, have an obvious influence on the fit of the

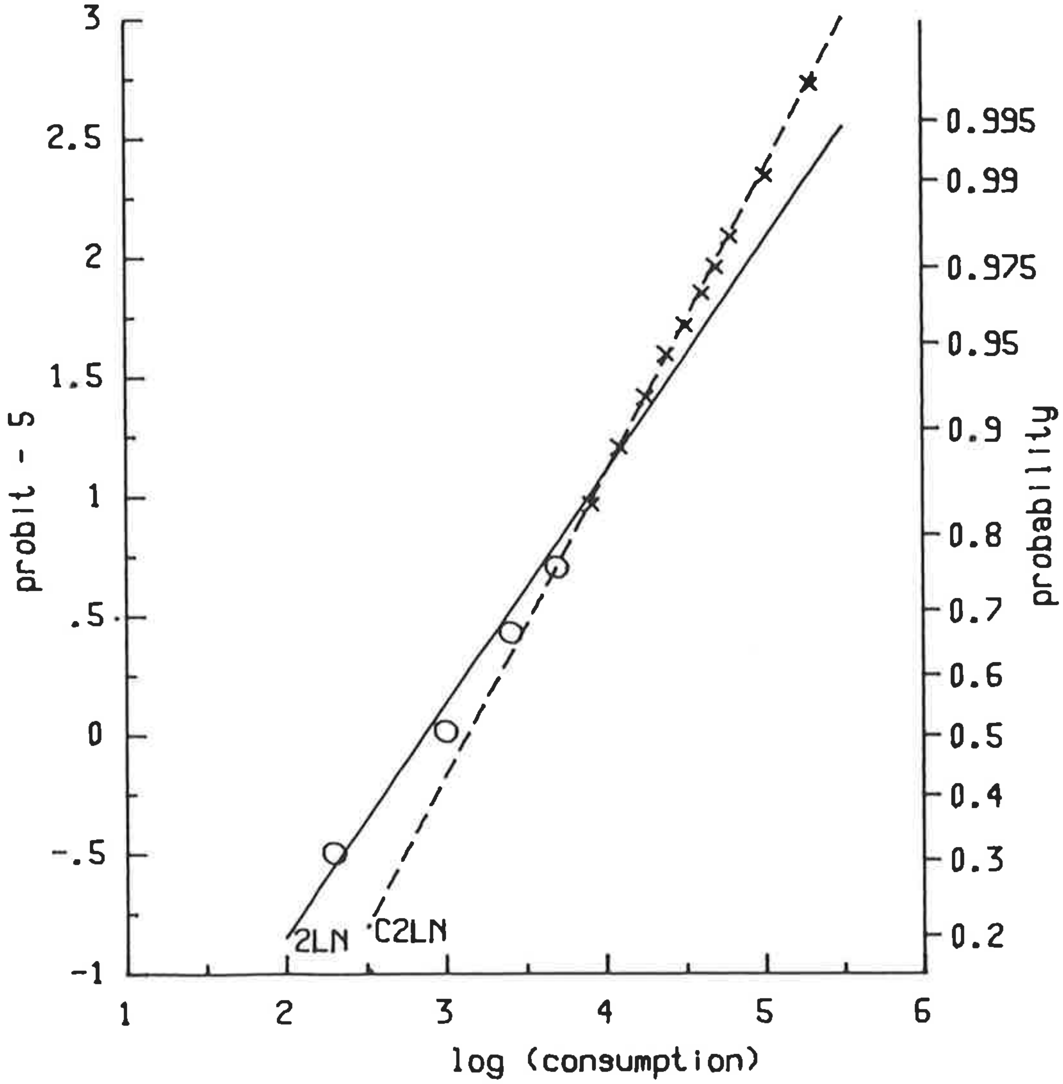

Figure 6.4 Logprobability plot for ABS male data.
(2LN = fitted two parameter lognormal distribution;
C2LN = fitted censored two parameter lognormal distribution)

distribution. Censoring the distribution at 50 g/day is equivalent to removing from the plot the four points shown as circles in Figure 6.4. The censored fit is represented by the dashed line. (χ^2 = 9.8, P = 0.28). Thus censoring the lower tail has improved the fit in the upper tail. The Appendix tables show that in all cases, with appropriate choice of the point of censorship, the data for adult males and females does not deviate significantly from a censored two parameter lognormal distribution. A truncated two parameter lognormal distribution also gave an adequate fit for those data sets for which it was fitted. (ABS males and females, Tables A17, A18; Busselton males, 1975 and 1978, Tables A8 and A20).

A note on terminology: A distribution which is neither censored nor truncated we shall term "unrestricted" where it is necessary to distinguish it from the other cases.

Other distributions which were tried on some of these data sets were a censored gamma distribution, with moderate success (see, for example, Table A7), and several with little or no success: exponential (one parameter), Weibull (two parameters), beta type II (three parameters) and log-hypergeometric (four parameters).

6.4.3 Fits to age subgroupings We return now to the problem of the age subgroupings. In fitting a specification to a subpopulation, we quite often find we are dealing with only a small sample size. We have already said that the median size for an age x sex subpopulation in these data sets is 125. When it is considered that these observations have a positively skewed distribution over several class intervals of alcohol consumption, it is not surprising that some of the frequencies in the upper tail are very small.

However for completeness, and in recognition of the fact that in some instances interest will lie in a particular agegroup rather than in the total population, Tables A2 to A24 list details of fits to individual agegroups in many instances.

While we would expect that there may be differences between agegroups in consumption patterns, it may be reasonable to assume that these differences vary continuously over the age range of the population, without any discontinuity. The model given in Section 4.5 allows us to assume a lognormal distribution for each subpopulation, with the parameters having a quadratic relation with age. Thus if t is the age in years, we can assume

$$\mu = a_0 + a_1 t + a_2 t^2$$

$$\sigma = b_0 + b_1 t + b_2 t^2$$

and estimate the coefficients a_i and b_i from all the subpopulation data, always providing there are enough agegroups for the quadratic fit. Substituting an appropriate value of t gives values for μ and σ defining a lognormal distribution for age t.

This model was fitted to several of the data sets (Tables A4, A8, A9, A13, A14, A15, A17, A18, A20, A21, A22). Details given in these tables are the equations for μ and σ, the overall χ^2 value for the fit, plus the predicted fits at values of t corresponding to the actual age subgroups. The overall χ^2 value is the sum of the χ^2 values for the individual agegroups. In only two cases (Busselton 1975 females, Table A9; RAAF outgoing recruits, Table A22) was the overall χ^2 value non-significant; however in nine of the other ten cases, all subgroups but one gave non-significant χ^2 values.

Let us take the Busselton 1978 females (Table A21) as an example. The model was fitted using data from all agegroups except 50-59 years, as this agegroup contained only three class intervals. Note that the remaining agegroups have differing numbers of class intervals, but all have a minimum of four (Table 5.24). The fitted model gives

$$\mu = 0.7298 + 0.0564\,t - 0.000505\,t^2$$

$$\sigma = 0.9150 + 0.0023\,t - 0.000029t^2$$

(χ^2 = 16.99, P=0.030). Taking for example, t = 35, we find μ = 2.086 and σ = 0.961, which defines a two parameter lognormal model for age 35 years. A χ^2 test of the probabilities predicted by this model and the data for the agegroup 30-39 gives a value of χ^2 which is significant at P=0.046. However there are no significant differences between the predicted models and the data for any of the other agegroups, and we may be prepared to accept the overall model, and use it as a basis for prediction. Figure 6.5 shows the two fitted curves for μ and σ, and as a comparison, the values obtained by fitting a two parameter lognormal distribution directly to each age group.

As an example of a possible predictive use, taking t = 55, the model provides a parametric estimate of the distribution of consumption in the 50-59 year agegroup. This is not estimable directly from the data. As another example, we may have available data on the age distribution of the population in five year class intervals. We could estimate the proportion of heavy drinkers in each of these class intervals from the overall model, and combine this information with the population estimates to get an estimate of the total number of heavy drinkers in the population.

For the cases of the ABS males (Table A17) and Busselton 1975 males (Table A8) the model has been fitted using a censored two parameter

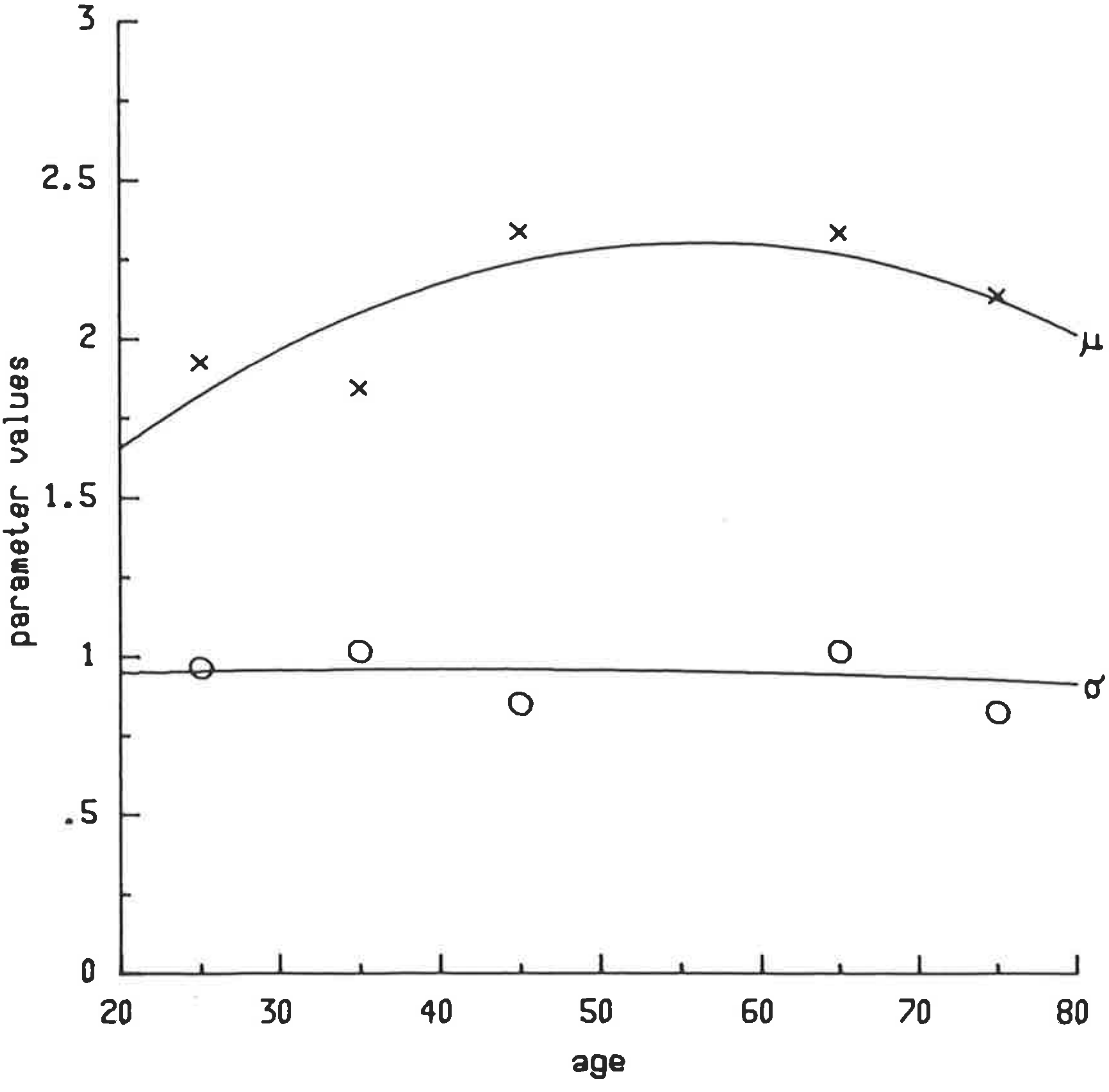

Figure 6.5 Parameter values for fits to overall and individual agegroup regressions for Busselton females, 1978.

lognormal distribution. For the RAAF recruits, only the linear term was used in the prediction of μ and σ.

6.4.4 The relation between the parameters of the two parameter lognormal fits Following Ledermann's (1956) work, it was assumed that there was some sort of constant relationship between the parameters of the lognormal distribution such that the proportion of heavy users is related to mean consumption (for example, Skog, 1982). While this was often stated, it has rarely been tested empirically. The current data presents an opportunity to look at this relationship.

On Figure 6.6 are plotted the values of μ and σ obtained from all the unrestricted two parametric lognormal fits which gave non-significant χ^2 values when fitted to aggregate adult and aggregate youth samples. The two sexes, and adult and youth samples, have been distinguished as shown in the key to the figure. Inspection of the figure shows that values of μ for adult populations typically lie between 2 and 4, and for σ, between 0.6 and 1.2. Youth populations tend to have lower values of μ, and higher values of σ, than adult ones. There is little evidence of a relationship between μ and σ in any one group, or overall. The one anomalous value for adult females (μ=0.5, σ=1.75) is from the Prahran survey (Tables 5.10 and A4). There appears to be no particular reason for this.

6.4.5 Comparison of censored and truncated lognormal fits Where the data have permitted, both censored and truncated lognormal distributions have been fitted. It has been said above that, in the context of grouped data, censoring information in the lower tail involves combining two or more of the lower class intervals, and fitting a lognormal distribution to the resulting fre-

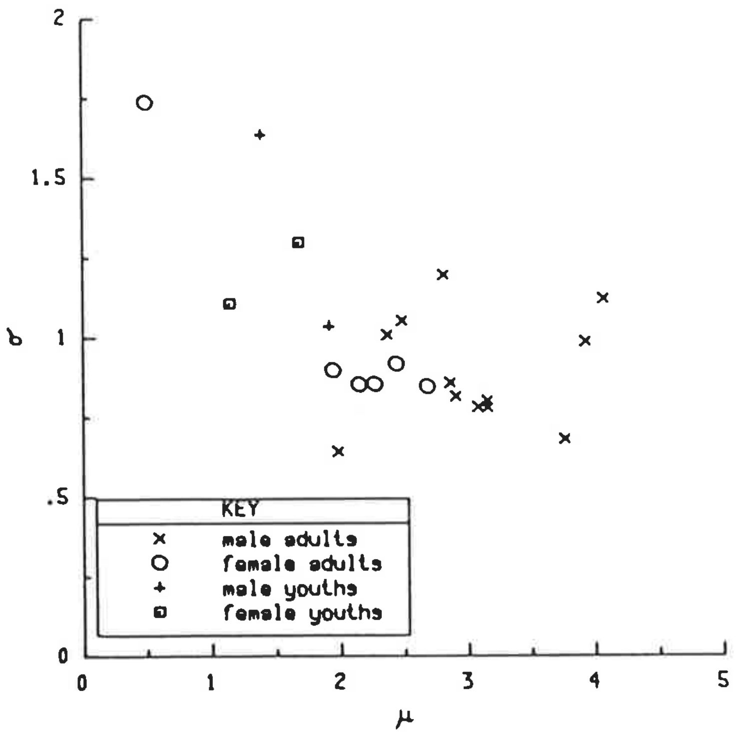

Figure 6.6 Parameters of fitted lognormal distributions.

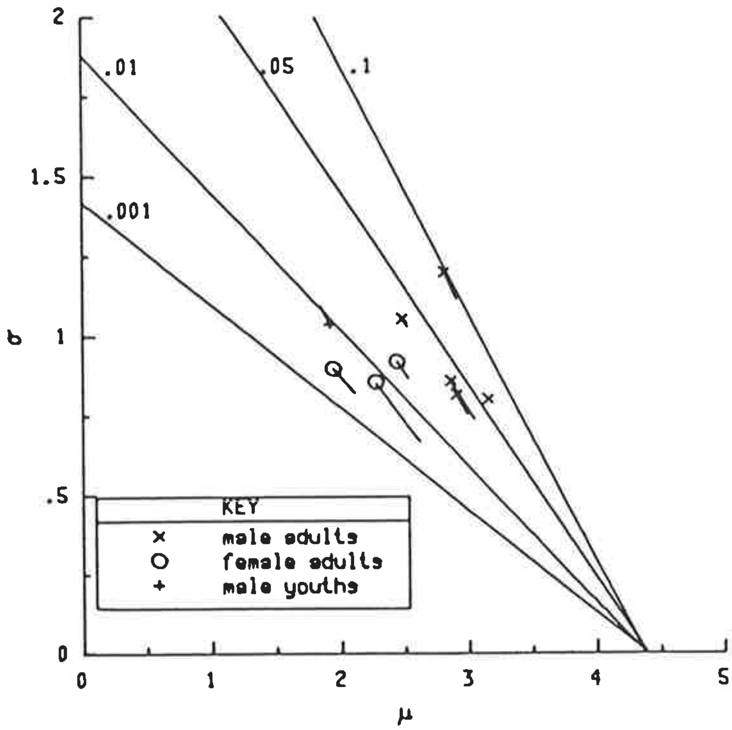

Figure 6.7 Parameters of fitted uncensored and censored lognormal distributions.

quencies. This is in contrast to truncation, where we delete some of the lower classes. We have supposed that both these procedures may result in a fit downgrading the effect of the observations in the lower tail, and thus depending more on the upper tail of the distribution.

One of the data sets (ABS males, Tables 5.21 and A17) had enough class intervals to enable several censored and truncated distributions to be fitted. Details are given in Table A17. An unrestricted lognormal distribution did not fit the data at all well (χ^2_{12} = 236), however for a point of censorship or point of truncation at or above 40 grammes per day, adequate fits to the data were obtained. The censored fit has two substantial advantages over the truncated fit however.

All censored fits use 100% of the available data, in marked contrast to the truncated fits. With a point of truncation at 40 grammes per day, only 24% of the data was used, and this percentage steadily declined: truncation at 60 g/day used only 11% of the data, and at 80 g/day, less than 6% was used.

Table 6.2

Comparison of censored and truncated distributions
Prediction of p(60) and p(80)

Data: ABS survey, males, all ages [Tables 5.21 and A17]

ζ	censored at ζ p(60)	censored at ζ p(80)	truncated at ζ p(60)	truncated at ζ p(80)
20	.111	.062	.154	.081
30	.111	.058	.207	.106
40	.113	.058	.129	.066
50	.115	.058	.062	.031
60	.113	.058	.045	.022
70	.110	.057	.101	.052
80	.105	.055	.381	.222

The second advantage of the censored fit is the stability of the upper tail produced by successive censorship of class intervals, compared with similar truncated fits. Table 6.2 shows predicted proportions of the tail above both 60 and 80 g/day. Over the whole range of points of censorship used, the predicted proportions of consumers drinking more than 60 and 80 g/day remained at approximately 0.11 and 0.06 respectively, although there is some variation. There is a slight reduction in the p(80) values with increasing censorship, and as the point of censorship increases the p(60) values initially increase, but then decrease for censorship at levels greater than 50 g/day. By contrast the similar proportions for the truncated distributions show no such stability.

6.4.6 <u>Comparison of censored and uncensored lognormal fits</u> We have already shown (Figure 6.4) the effect of fitting a censored rather than an uncensored distribution to one set of data, but in that case, the uncensored distribution did not give an acceptable fit to the data. It is of interest to examine the effect of censoring on those data sets for which both censored and uncensored lognormal distributions gave fits with non-significant χ^2 values. This is done in Figure 6.7, where the plotted points represent the parameters of the uncensored distribution. Male and female samples are represented with crosses and circles respectively. From each of these points a line has been drawn to the position of the parameters of the censored distribution. Superimposed on Figure 6.7 is the fan of contours of p(80), the proportion of consumers drinking in excess of 80 g/day. The contours are taken from Figure 4.8, and are given for p(80) = 0.1, 0.05, 0.01 and 0.001.

The effect of censoring the distribution is to move the values of μ and σ a small distance, almost parallel to the contours, but in most cases moving slightly in the direction of increased p(80). That is, if both the censored and uncensored fits give adequate fits to the one data set, the censored fit has usually predicted a slightly greater percentage of heavy users than has the uncensored fit. The notable exception on the figure is for the Geelong school survey males, the only youth sample included. By comparison of the changes shown in Figure 6.7 with the scatter associated with μ and σ shown in Figure 6.6, we note that censoring the distribution has caused small rather than gross changes in the values of the parameters.

6.5 Mean consumption and proportion of heavy consumers

The early applications of Ledermann's work assumed a mathematical relationship between the mean consumption and the proportion of excessive users (see Chapter 2). Subsequently, following the research which lead up to the report of Bruun *et al* (1975), this relationship was supposed to hold only approximately, and alcohol control policies have been based on an observed empirical relationship between mean consumption and excessive consumption. We now examine the nature of this relationship for the Australian data.

An immediate problem is what values of mean consumption and proportion of excessive users to use. There are two candidates:

i. non-parametric estimates obtained directly from the data

ii. parametric estimates obtained from distributions fitted to the data

Parametric estimates have the advantages of smoothing the grouped frequencies and being less sensitive to sampling errors. Additionally, estimates of tail probabilities may not be available directly from the data. Assuming that there exists a distribution which provides an "adequate" fit to the data, we prefer estimates based on it. Our approach has been to estimate the mean consumption from an unrestricted fit, and the proportions of heavy consumers from a censored fit, where these are both available (that is, there are enough class intervals to fit both forms of the distribution, and both have χ^2 values which are non-significant at P = 0.05), since in estimating the mean we are interested in using all the data, while in estimating the proportion of heavy consumers we require a mode of inference which concentrates on the upper tail. Where both two and three parameter lognormal fits are available, estimates from the two parameter fit have been used, since in general they

Table 6.3

"Best estimate" of proportions of consumers with consumption in excess of 60 and 80 g/day
(Samples are adults unless otherwise stated)

Sample	source of estimates mean/heavy consumers	mean (g/day)	p(60)	p(80)
	Females			
Prahran	3LN/C2LN	14.0	.019	.013
Busselton 1975	2LN/C2LN	17.5	.035	.016
AWU members	NP/NP	10.1	-	.003
Medicheck 1975	NP/NP	16.9	-	.006
Medicheck 1976	NP/NP	15.4	-	.004
Geelong beach (youth)	2LN/2LN	12.8	.030	.017
Geelong beach	NP/NP	21.6	.019	-
NW Melbourne	NP/NP	9.3	-	.001
Geelong school (youth)	2LN/C2LN	6.7	.004	.002
ABS	2LN/C2LN	11.2	.011	.004
Busselton 1978	2LN/C2LN	13.9	.014	.004
Qld. h-s students	2LN/C2LN	10.5	.008	.003
Townsville	2LN/2LN	20.9	.048	.022
	Males			
Perth 1965-66	2LN/2LN	109.5	.490	.389
Perth 1978-79	2LN/2LN	82.5	.433	.322
Prahran	2LN/C2LN	34.2	.144	.093
Busselton 1975	2LN/C2LN	25.6	.072	.033
AWU members	2LN/2LN	17.8	.044	.023
Medicheck 1975	NP/NP	26.7	-	.048
Medicheck 1976	NP/NP	25.8	-	.041
Geelong beach (youth)	NP/NP	34.3	.193	-
Geelong beach	NP/NP	42.9	.331	-
NW Melbourne (youth)	2LN/2LN	15.4	.049	.034
NW Melbourne	NP/NP	25.9	-	.072
Geelong school (youth)	2LN/C2LN	11.7	.020	.010
ABS	NP/C2LN	28.6	.115	.058
Adelaide Uni.	2LN/C2LN	9.0	.003	.001
Busselton 1978	3LN/C2LN	26.3	.098	.048
RAAF incoming	2LN/C2LN	25.4	.078	.035
RAAF outgoing	2LN/C2LN	32.4	.121	.063
Qld h-s students	2LN/C2LN	21.0	.063	.035
Townsville	2LN/2LN	54.2	.312	.181
	Mixed			
SE SA	NP/NP	10.2	-	.008
Road crash victims	2LN/C2LN	25.6	-	.076

will have lower standard errors than those based on the three parameter fit.

Table 6.3 lists the estimates, together with their sources. The following abbreviations are used for the sources of the estimates:

2LN: two parameter lognormal distribution

3LN: three parameter lognormal distribution

C2LN: censored two parameter lognormal distribution

NP: non-parametric

The relationship of the proportion of females drinking more than 60 g/day [p(60)] and more than 80 g/day [p(80)] to the mean consumption is illustrated in Figures 6.8 and 6.9 respectively. Parametric estimates are shown in circles, while non-parametric ones are shown as crosses. Figures 6.10 and 6.11 present similar data for males, but with very different scales to the previous two figures. Because of these scale differences the data for both sexes is combined in Figures 6.12 and 6.13.

We note firstly that there is no consistent difference between the parametric and non-parametric estimates. This can be further checked by comparing the parametric estimates of Table 6.3 with non-parametric ones derived from the data given in Chapter 5, and again there is no consistent pattern of differences.

There is an obvious relationship between proportion of heavy users and mean consumption. Over the ranges of data present for females, the relationship is linear, and described by the regressions (letting x = mean consumption in g/day)

$$p(60) = -0.0097 + 0.002132\,x$$

standard errors: 0.0113, 0.000749; Fisher's A = 47.0%; n = 9

$$p(80) = -0.0092 + 0.001290\,x$$

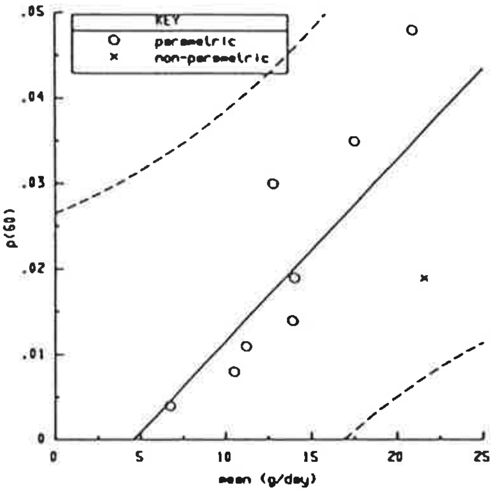

Figure 6.8 'Best estimate' of p(60) - females.

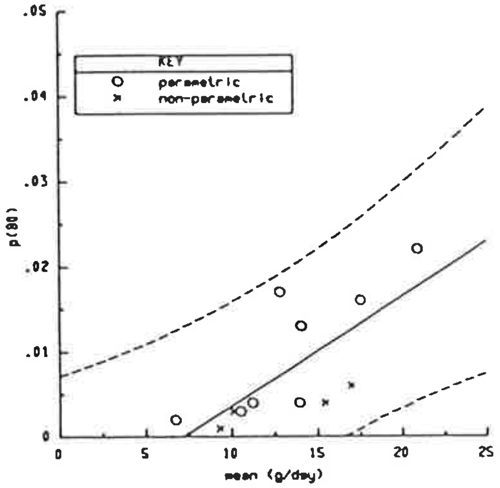

Figure 6.9 'Best estimate' of p(80) - females.

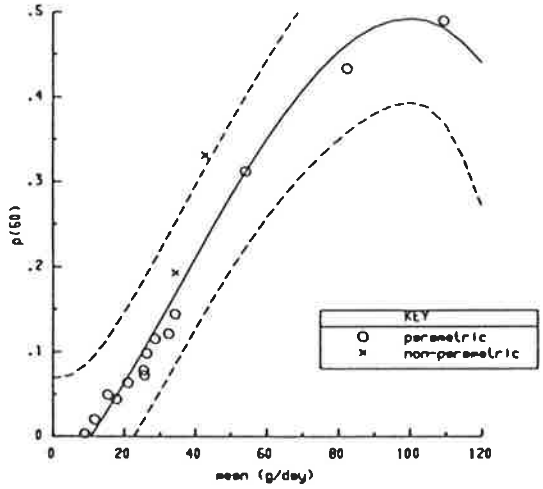

Figure 6.10 'Best estimate' of p(60) - males.

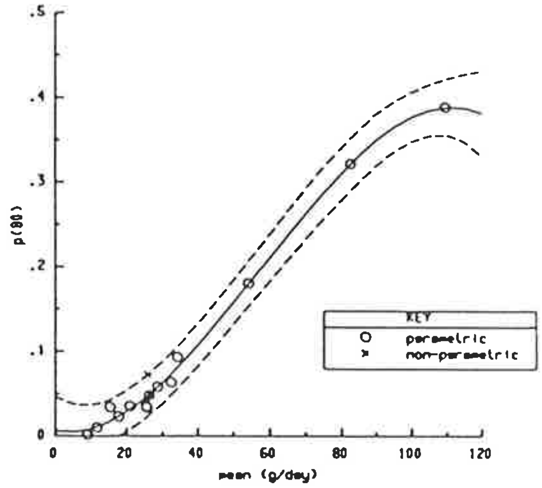

Figure 6.11 'Best estimate' of p(80) - males.

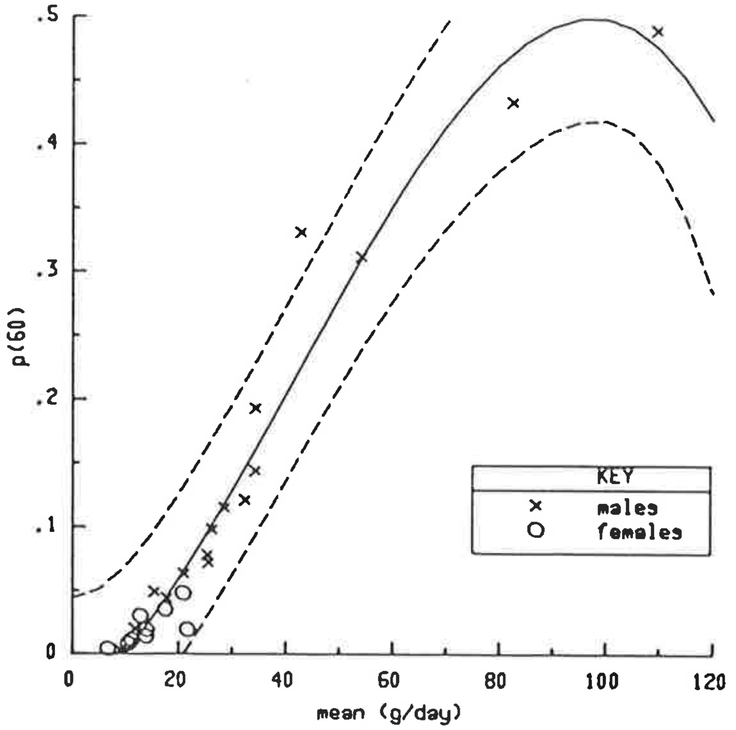

Figure 6.12 'Best estimate' of p(60) - both sexes.

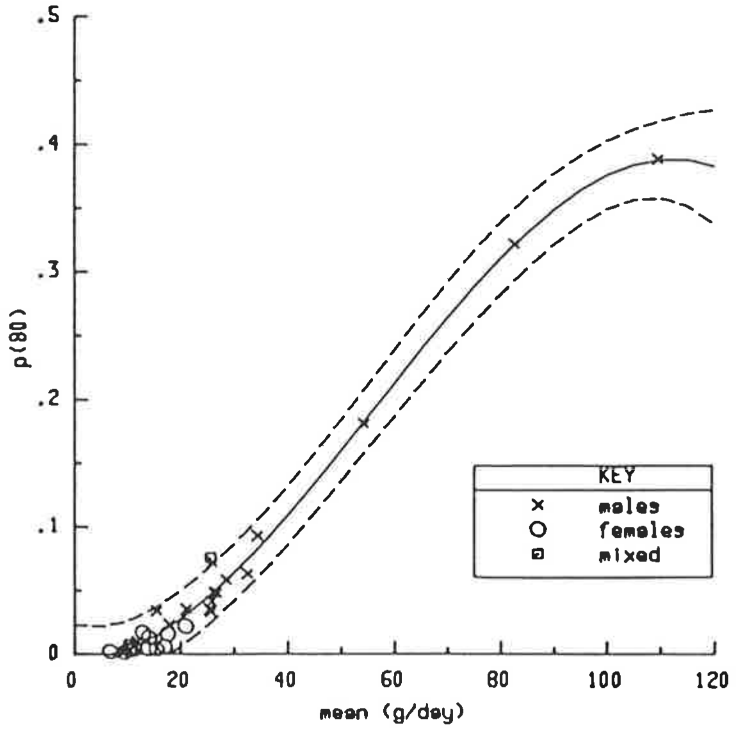

Figure 6.13 'Best estimate' of p(80) - both sexes.

standard errors: 0.00531; 0.000385; Fisher's A = 48.3%; n = 12

These regressions are plotted on Figures 6.8 and 6.9, together with a 95% confidence interval for one extra observation. There was no significant evidence of curvilinear behaviour.

The data for males however extends over a much greater range of mean values and there is evidence of slight curvature in the relationships. Cubic polynomials were fitted to both p(60) and p(80) data, with the following results

$$p(60) = -0.0623 + 0.00529\,x + 0.0000618\,x^2 - 0.000000592\,x^3$$

standard errors: 0.0481, 0.00360, 0.0000760, 0.000000448;

Fisher's A = 94.2%; n = 16

$$p(80) = 0.0058 - 0.00068\,x + 0.0001053\,x^2 - 0.000000613\,x^3$$

standard errors: 0.0147, 0.00108, 0.0000224, 0.000000132;

Fisher's A = 99.0%; n = 17

These are plotted on Figures 6.10 and 6.11, together with a 95% confidence interval.

Similar regressions describe the data from both sexes together, and these are plotted on Figures 6.12 and 6.13.

$$p(60) = -0.0414 + 0.00331\,x + 0.0001035\,x^2 - 0.000000826\,x^3$$

standard errors: 0.0275, 0.00239, 0.0000547, 0.000000336;

Fisher's A = 95.0%; n = 25

$$p(80) = -0.00476 - 0.000032\,x + 0.0000950\,x^2 - 0.000000565\,x^3$$

standard errors: 0.00820, 0.000733, 0.0000173, 0.000000108;

Fisher's A = 98.6%; n = 31

While these regressions include a cubic term, in practical terms, over the usual range of mean values, say up to 50 g/day, the relationship is essentially quadratic.

6.6 Discussion

Over the last thirty years, there has been interest in fitting statistical distributions to alcohol consumption data. While the main interest has revolved around the use of the two parameter lognormal distribution, other distributions, such as the gamma distribution, have been considered. But in all the attempts, there has seldom been any explicit recognition of the inference to be made from use of a particular specification. This has resulted in various conflicts of interest.

Despite a declining interest in recent years in the explicit fitting of distributions, it is still useful in a variety of situations, among the most important being to judge the effect of alcohol control policies on heavy consumption.

In this chapter we have shown that when our main interest lies in inference about the upper tail of the distribution, the conventional unrestricted two parameter distribution may not be a suitable choice of model. This is because, on the logarithmic scale, the distribution demands a strict symmetry, and so the shape of the upper tail is constrained to correspond to the shape of the lower tail. We have suggested that better models for this purpose may result from censoring the lower tail of the distribution, or from adding a third parameter to the distribution, both of which measures we would expect, heuristically, to leave the observations on heavy drinkers more free to determine the upper tail.

We have shown that for those data sets containing enough class intervals, the censored two parameter lognormal distribution gives a good fit to the upper tail, and have used these facts to examine the relationship between mean consumption and prevalence of heavy drinkers, showing that over the common range of mean consumptions, the relationship is approximately quadratic.

However the fact that there is an empirical relationship between proportion of heavy consumers and mean consumption derived from about twenty Australian surveys is not necessarily applicable directly to changes over time within the one population.

The data considered does however contain some information on longitudinal changes. Both the Busselton surveys (1975 and 1978) and the Medicheck surveys (1975 and 1976) contain information for both sexes, and additionally the surveys of Perth social drinkers (1965-66 and 1978-79) and the RAAF recruits (ingoing and outgoing) contain information on male drinkers. Details of mean consumption and proportion of heavy users for these surveys are contained in Table 6.3. In all cases, the changes in mean consumption and proportions of heavy use are in in the same direction, four showing decreasing consumption and two surveys showing increasing consumption. Thus we do have some evidence that, under the conditions prevailing in these particular populations, mean consumption and proportion of heavy consumers both increase and decrease together.

But it is important to remember that values of both mean consumption and proportion of heavy drinkers are the results of drinking habits of the population, and are not variables which can be directly changed. Any change in the drinking habits of the population will be reflected by changes in both these statistics.

Without a thorough examination of the conditions under which these changes in drinking habits occurred, it is not possible to infer directly from these results to public health policies aimed, say, at reducing mean consumption in the hope that excess consumption will also decline. To do this, it would be necessary to know that the observed changes have resulted, via changes in drinking habits, from such things as, for example, price increases and reduced availability of beverage outlets, rather than representing random fluctuations in drinking habits.

Appendix

Details of fitted distributions

Table A1

Sample statistics – consumers

Sample	sample size	no of classes	mean	s.d. (log)	skewness
Heyfield [Table 5.7]					
male youths (14-21)	37	3	11.7	1.19	1.84
males 22-64	402	3	23.0	1.29	0.38
males 65+	28	3	11.2	0.88	2.07
male adults (22+)	430	3	22.2	1.29	0.46
female youths (14-21)	24	1	2.7	–	–
females 22-64	241	3	8.6	0.79	2.69
females 65+	17	3	23.5	1.58	0.18
female adults (22+)	258	3	9.5	0.91	2.44
Perth social drinkers [Table 5.8]					
males 1965-66	287	4	77.3	1.23	0.56
males 1978-79	253	4	68.7	1.19	0.87
Alcoholism clinic [Table 5.9]					
males	143	10	226.7	0.40	1.15
females	28	6	166.1	0.32	1.33
Prahran [Table 5.10]					
male youths (10-19)	42	4	11.1	0.63	5.06
males 20-29	201	4	30.3	1.13	1.43
males 30-39	119	4	31.1	1.15	1.41
males 40-49	90	4	32.5	1.10	1.50
males 50-59	76	4	27.1	1.09	1.65
males 60-69	66	4	37.7	1.27	0.99
males 70+	40	4	22.3	1.04	2.00
male adults (20+)	592	5	32.4	1.14	1.90
female youths (10-19)	63	2	6.2	0.39	29.98
females 20-29	217	4	11.1	0.62	4.94
females 30-39	84	4	16.8	0.95	3.05
females 40-49	70	4	15.9	0.87	3.13
females 50-59	88	3	15.2	0.86	1.79
females 60-69	56	4	13.4	0.78	4.02
females 70+	55	2	7.2	0.73	3.63
female adults (20+)	570	5	13.5	0.76	4.70

Table A1
(continued)

Sample statistics – consumers

Sample	sample size	no of classes	mean	s.d. (log)	skewness
Bourke aborigines [Table 5.11]					
male youths (10-19)	16	3	16.3	1.08	1.78
males 20-29	48	6	98.4	1.20	-0.18
males 30-39	35	6	90.6	1.52	0.15
males 40-49	15	6	134.3	0.59	-0.33
males 50+	18	6	117.2	1.14	-0.03
male adults (20+)	109	6	100.6	1.29	0.08
females 20-29	13	3	27.3	1.05	0.71
females 30-39	6	5	45.8	1.72	1.30
females 40-49	9	4	26.7	1.10	2.63
females 50+	6	5	53.3	1.82	0.93
female adults (20+)	34	5	35.4	1.24	1.85
South East of SA [Table 5.12]					
all persons	523	4	10.2	0.58	5.65
Road crash victims [Table 5.13]					
BAC < 0.049	134	6	15.7	1.27	3.52
BAC > 0.05	77	6	42.9	1.07	1.26
all	211	6	25.6	1.41	2.08
Busselton, WA, 1975 [Table 5.14]					
males < 30	234	6	21.7	0.79	2.26
males 30-39	203	6	24.5	0.88	2.25
males 40-49	190	6	24.0	0.85	2.41
males 50-59	199	6	25.8	0.91	1.94
males 60-69	178	6	27.1	0.91	1.91
males 70+	121	6	26.9	0.95	1.90
male adults (<30+)	1135	6	25.1	0.89	2.03
females < 30	154	6	16.0	0.35	4.35
females 30-39	129	6	17.1	0.48	4.20
females 40-49	147	6	17.6	0.53	3.79
females 50-59	148	4	18.6	0.64	2.30
females 60-69	109	6	23.0	0.84	2.15
females 70+	53	6	20.6	0.71	3.09
female adults (<30+)	740	6	18.4	0.60	3.35
AWU members [Table 5.15]					
males	2662	4	19.8	0.95	2.60
females	5074	4	10.1	0.53	4.51

Table A1
(continued)

Sample statistics – consumers

Sample	sample size	no of classes	mean	s.d. (log)	skewness
Medicheck screenings [Table 5.16]					
males 1975	3653	4	26.7	0.93	2.17
males 1976	5956	4	25.8	0.95	2.14
females 1975	1816	4	16.9	0.82	2.90
females 1976	3446	4	15.4	0.79	2.80
Geelong beach survey [Table 5.17]					
males 10–14	28	2	19.2	0.71	11.85
males 15–19	355	4	35.5	0.93	1.34
males 20–24	105	4	45.6	1.09	0.52
males 25+	55	4	37.9	1.00	1.22
male youths (10–19)	383	4	34.3	0.90	1.45
male adults (20+)	160	4	42.9	1.07	0.74
females 10–14	28	2	18.2	0.43	56.48
females 15–19	256	4	19.2	–	7.70
females 20–24	74	3	23.5	0.39	2.84
females 25+	30	2	17.0	–	–
female youths (10–19)	284	4	19.1	–	8.15
female adults (20+)	104	3	21.6	–	4.11
North West Melbourne [Table 5.18]					
male youths (15–19)	91	4	15.1	0.90	2.82
males 20–29	208	4	21.7	1.13	1.81
males 30–39	172	4	25.0	1.18	1.56
males 40–49	155	4	33.7	1.35	0.94
males 50–59	117	4	25.3	1.17	1.58
males 60–69	78	4	27.2	1.25	1.35
males 70+	32	4	19.2	1.07	2.15
male adults (20+)	762	4	25.9	1.21	1.46
female youths (15–19)	85	2	5.6	–	–
females 20–29	200	4	8.8	0.37	5.10
females 30–39	149	3	8.9	0.43	3.92
females 40–49	129	3	10.7	0.75	2.90
females 50–59	106	3	9.1	0.49	3.94
females 60–69	60	3	9.5	0.52	3.55
females 70+	39	2	6.2	0.41	26.87
female adults (20+)	683	4	9.3	0.45	4.34

Table A1
(continued)

Sample statistics – consumers

Sample	**sample size**	**no of classes**	**mean**	**s.d. (log)**	**skewness**
Geelong school survey [Table 5.19]					
males 14	307	6	9.3	0.71	4.30
males 15	496	6	12.3	0.89	3.32
males 16	371	6	13.4	0.90	3.35
males 17	182	6	13.8	0.90	3.09
males 18	52	5	15.7	0.85	2.53
male youths (all)	1424	6	12.3	0.87	3.40
females 14	233	6	6.9	0.44	7.30
females 15	386	5	7.4	0.53	5.67
females 16	292	5	7.6	0.55	5.83
females 17	178	4	8.2	0.64	2.59
females 18	24	2	5.3	0.52	9.40
female youths (all)	1127	6	7.5	0.54	5.82
Newcastle alcohol abusers [Table 5.20]					
all persons	100	13	233.5	0.42	2.66
ABS survey [Table 5.21]					
males 18-24	1064	9	26.9	1.09	1.12
males 25-44	2618	15	29.5	1.11	2.39
males 45-64	1753	11	28.5	1.11	1.33
males 65+	464	9	17.8	1.00	1.83
male adults (18+)	5897	15	28.6	1.12	2.55
females 18-24	757	7	10.5	0.68	2.99
females 25-44	1765	7	11.2	0.73	2.78
females 45-64	1100	7	12.5	0.81	2.26
females 65+	345	7	9.6	0.60	3.61
female adults (18+)	3966	7	11.3	0.74	2.70
Sydney Hospital Health Service [Table 5.22]					
males	10474	3	22.1	1.10	0.20
females	3962	3	13.7	0.83	1.14
Adelaide University students [Table 5.23]					
males	108	5	8.9	0.52	3.23
females	46	3	7.0	0.12	4.81

Table A1
(continued)

Sample statistics – consumers

Sample	sample size	no of classes	mean	s.d. (log)	skewness
Busselton, WA, 1978 [Table 5.24]					
males <30	249	6	28.6	0.95	1.88
males 30-39	237	6	25.2	0.89	1.97
males 40-49	205	6	27.3	0.95	1.90
males 50-59	210	6	31.9	1.04	1.26
males 60-69	202	6	28.3	0.99	1.57
males 70+	158	6	22.9	0.83	2.51
male adults (<30+)	1260	6	27.5	0.95	1.78
females < 30	201	4	13.6	–	6.13
females 30-39	203	5	13.6	–	6.03
females 40-49	168	4	16.0	0.37	3.36
females 50-59	181	3	16.9	0.53	2.29
females 60-69	168	6	18.5	0.60	3.29
females 70+	80	4	13.8	–	6.02
female adults (<30+)	1001	6	15.4	0.25	4.25
RAAF recruits [Table 5.25]					
incoming 17-20	260	5	23.3	0.30	3.37
incoming 21-25	107	5	30.2	0.79	1.89
incoming 26+	70	5	27.1	0.62	2.31
incoming adults (all)	444	5	25.6	0.54	2.65
outgoing 17-20	286	5	29.7	0.76	1.96
outgoing 21-25	121	5	30.7	0.81	1.75
outgoing 26+	73	5	34.1	0.91	1.40
outgoing adults (all)	491	5	30.6	0.80	1.81
Queensland human-service students [Table 5.26]					
males	684	5	20.9	0.75	2.36
females	583	5	13.2	–	7.54
Townsville residents [Table 5.27]					
male youths (15-17)	6	2	26.6	1.06	7.20
males 18-24	35	4	53.1	0.97	1.23
males 25-44	44	4	56.4	1.08	0.74
males 45-46	19	4	43.2	0.92	1.78
males 65+	12	3	43.3	0.95	1.00
male adults (18+)	110	4	51.6	1.01	1.07
female youths (15-17)	2	1	20.0	–	–
females 18-24	23	4	27.0	–	5.79
females 25-44	26	2	23.0	–	–
females 45-64	18	2	28.8	1.16	3.75
females 65+	9	2	24.4	0.61	27.04
female adults (18+)	76	4	25.8	–	9.01

Table A2

Details of fitted distributions

Perth "social drinkers" [Table 5.8]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males 1965-66	287	4.067, 1.121	0.28	1	.594
males 1978-79	253	3.927, 0.986	0.81	1	.367
gamma (α,β)					
males 1965-66	287	0.964, 92.937	3.02	1	.082
males 1978-79	253	1.104, 66.445	4.51	1	.034

Table A3

Details of fitted distributions

Alcoholism Clinic [Table 5.9]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males	143	5.344, 0.404	33.81	6	<.001
females	28	5.056, 0.333	19.18	2	<.001
two parameter lognormal, censored below 200 g/day					
males	143	5.366, 0.393	8.78	5	.118
females	28	5.098, 0.361	0.57	1	.450
three parameter lognormal (μ,σ,τ)					
males	143	4.788, 0.685, 78.951	25.94	5	<.001
gamma (α,β)					
males	143	6.142, 37.010	42.42	7	<.001
females	28	8.685, 19.164	21.83	3	<.001

Table A4

Details of fitted distributions

Prahran [Table 5.10]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males 20-29	201	2.817, 1.190	1.43	1	.232
males 30-39	119	2.834, 1.236	0.17	1	.680
males 40-49	90	2.950, 1.152	1.65	1	.199
males 50-59	76	2.701, 1.119	1.13	1	.289
males 60-69	66	3.017, 1.605	1.78	1	.183
males 70+	40	2.458, 1.091	0.76	1	.382
male adults (20+)	592	2.816, 1.198	2.71	2	.258
females 20-29	217	1.668, 0.990	4.75	1	.029
females 30-39	84	1.646, 1.587	5.29	1	.021
females 40-49	70	2.078, 1.031	0.03	1	.856
females 60-69	56	1.770, 1.130	0.81	1	.367
female adults (20+)	570	1.800, 1.090	7.46	2	.024
two parameter lognormal, censored below 40 g/day					
male adults (20+)	592	2.914, 1.110	1.61	1	.205
female adults (20+)	570	0.506, 1.737	2.11	1	.146
three parameter lognormal (μ,σ,τ)					
male adults (20+)	592	3.063, 1.035, -3.877	1.30	1	.253
female adults (20+)	570	-0.144, 1.975, 7.875	2.48	1	.115
gamma (α,β)					
males 20-29	201	0.780, 37.994	0.26	1	.610
males 30-39	119	0.740, 42.052	0.77	1	.380
males 40-49	90	0.859, 38.153	5.84	1	.016
males 50-59	76	0.816, 31.066	0.00	1	>.999
males 60-69	66	0.543, 86.655	4.54	1	.033
males 70+	40	0.754, 26.261	0.04	1	.842
male adults (20+)	592	0.723, 42.974	7.61	2	.023
females 20-29	217	0.490, 16.173	8.90	1	.003
females 30-39	84	0.289, 52.329	8.86	1	.047
females 40-49	70	0.647, 20.060	0.75	1	.387
females 60-69	56	0.455, 22.779	2.06	1	.151
female adults (20+)	570	0.464, 22.437	21.83	2	<.001

Table A4
(continued)

Details of fitted distributions

Prahran [Table 5.10]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal with covariance on t = age in years					
male adults (20+)	592	$\mu = 2.1954 + .0331t - .00038t^2$			
		$\sigma = 1.0019 + .0083t - .00006t^2$	17.70	6	.007
t=25		2.784, 1.168	1.66	1	.197
t=35		2.887, 1.213	0.36	1	.546
t=45		2.912, 1.244	2.16	1	.142
t=55		2.862, 1.263	3.66	1	.055
t=65		2.735, 1.269	8.09	1	.005
t=75		2.532, 1.262	1.76	1	.185

Table A5

Details of fitted distributions

Bourke aborigines [Table 5.11]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males 20-29	41	4.241, 1.058	37.77	3	<.001
males 30-39	35	3.949, 1.511	27.59	3	<.001
males 40-49	15	4.831, 0.548	3.58	2	.167
males 50+	18	4.585, 1.259	1.73	3	.630
male adults (20+)	109	4.248, 1.241	53.89	3	<.001
female adults (20+)	34	2.821, 1.398	0.47	2	.790
two parameter lognormal, censored below ζ g/day					
male adults (20+)	109	4.563, 0.585 (ζ = 80)	5.16	1	.023
female adults (20+)	34	2.768, 1.452 (ζ = 40)	0.46	1	.499
gamma (α,β)					
males 20-29	48	1.422, 69.492	23.32	3	<.001
males 30-39	35	0.765, 126.103	17.32	3	<.001
males 40-49	15	4.280, 32.425	2.27	3	.518
males 50+	18	1.126, 129.366	0.65	3	.885
male adults (20+)	109	1.104, 97.561	28.18	3	<.001
female adults (20+)	34	0.571, 64.185	1.76	2	.414
gamma, censored below ζ g/day					
male adults (20+)	109	2.839, 38.595 (ζ = 80)	3.07	1	.080
female adults (20+)	34	0.338, 105.820 (ζ = 40)	0.79	1	.374

Table A6

Details of fitted distributions

South East of South Australia [Table 5.12]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
all persons	523	1.674, 0.957	7.95	1	.005
gamma (α,β)					
all persons	523	0.546, 13.900	20.10	1	<.001

Table A7

Details of fitted distributions

Road crash victims [Table 5.13]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
BAC < .049	134	1.368, 1.798	7.26	3	.064
BAC > .05	77	3.301, 1.008	4.24	2	.120
all	211	2.088, 1.774	22.85	3	<.001
two parameter lognormal, censored below 40 g/day					
BAC > .05	77	3.473, 0.812	1.82	1	.177
all	211	2.739, 1.087	1.95	1	.163
three parameter lognormal (μ,σ,τ)					
BAC < .049	134	2.067, 1.213, -2.436	0.51	2	.774
all	211	2.727, 1.142, -3.840	2.62	2	.270
gamma (α,β)					
BAC < .049	134	0.422, 32.765	3.75	3	.290
BAC > .05	77	1.165, 35.765	0.79	3	.148
all	211	0.510, 47.237	0.96	3	.811
gamma, censored below 40 g/day					
BAC < .049	134	0.128, 78.493	0.02	1	.888
BAC > .05	77	1.160, 35.881	0.79	1	.374
all	211	0.447, 52.411	0.71	1	.399

Table A8

Details of fitted distributions

Busselton 1975, males [Table 5.14]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages < 30	234	2.708, 0.854	6.13	3	.105
ages 30-39	203	2.863, 0.843	2.58	3	.461
ages 40-49	190	2.926, 0.738	10.94	3	.012
ages 50-59	199	2.959, 0.791	5.38	3	.146
ages 60-69	178	3.065, 0.715	1.32	3	.724
ages 70+	121	2.891, 0.928	2.24	3	.525
adults (<30+)	1135	2.911, 0.816	3.58	3	.311
two parameter lognormal with covariance on t = age in years					
adults (<30+)	1135	$\mu = 2.0864 + .0309t - .00026t^2$			
		$\sigma = 1.3213 - .0227t + .00022t^2$	32.40	18	.020
t=25		2.695, 0.894	6.42	3	.093
t=35		2.848, 0.801	3.26	3	.353
t=45		2.948, 0.753	11.22	3	.011
t=55		2.997, 0.750	5.87	3	.118
t=65		2.993, 0.791	3.09	3	.377
t=75		2.937, 0.877	2.54	3	.468
two parameter lognormal, censored below 40 g/day					
ages < 30	234	2.981, 0.670	3.05	2	.218
ages 30-39	203	2.840, 0.860	2.56	2	.278
ages 40-49	190	2.974, 0.703	10.84	2	.004
ages 50-59	199	3.101, 0.683	4.03	2	.133
ages 60-69	178	3.054, 0.723	1.32	2	.518
ages 70+	121	2.861, 0.951	2.21	2	.330
adults (<30+)	1135	2.995, 0.753	1.51	2	.469
two parameter lognormal, censored below 40 g/day, with covariance on t					
adults (<30+)	1135	$\mu = 2.4808 + .0208t - .00020t^2$			
		$\sigma = 1.0474 - .0158t + .00018t^2$	27.35	12	.007
t=25		2.877, 0.766	3.79	2	.151
t=35		2.967, 0.717	4.05	2	.132
t=45		3.017, 0.704	11.19	2	.004
t=55		3.028, 0.727	4.35	2	.113
t=65		2.999, 0.786	1.60	2	.449
t=75		2.931, 0.881	2.37	2	.306

Table A8
(continued)

Details of fitted distributions

Busselton 1975, males [Table 5.14]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal, truncated below 20 g/day					
ages < 30	85	3.424, 0.544	3.01	2	.223
ages 30-39	89	2.621, 0.930	2.51	2	.286
ages 40-49	88	2.799, 0.785	10.91	2	.004
ages 50-59	95	3.332, 0.623	4.41	2	.110
ages 60-69	96	2.981, 0.749	1.29	2	.524
ages 70+	55	2.959, 0.901	2.23	2	.328
adults (<30+)	518	3.181, 0.703	1.89	2	.388
three parameter lognormal (μ,σ,τ)					
ages < 30	234	4.093, 0.363, -48.039	2.96	2	.227
ages 30-39	203	2.684, 0.929, 3.091	2.50	2	.287
ages 40-49	190	2.758, 0.813, 3.020	10.85	2	.004
ages 50-59	199	3.452, 0.581, -12.606	4.59	2	.101
ages 60-69	178	2.965, 0.761, 2.010	1.29	2	.526
ages 70+	121	2.991, 0.878, -2.019	2.22	2	.329
adults (<30+)	1135	3.218, 0.679, -6.911	2.04	2	.360
gamma (α,β)					
ages < 30	234	0.949, 21.093	3.20	3	.362
ages 30-39	203	1.054, 22.321	5.40	3	.145
ages 40-49	190	1.392, 16.824	14.35	3	.003
ages 50-59	199	1.257, 19.956	5.10	3	.165
ages 60-69	178	1.588, 16.866	3.80	3	.284
ages 70+	121	0.905, 28.555	2.69	3	.442
adults (<30+)	1135	1.154, 21.026	6.43	3	.093
gamma, censored below 40 g/day					
ages < 30	234	1.144, 18.532	2.86	2	.239
ages 30-39	203	0.691, 30.294	3.57	2	.168
ages 40-49	190	1.028, 20.921	13.56	2	.001
ages 50-59	199	1.218, 20.416	5.09	2	.079
ages 60-69	178	1.070, 22.645	1.98	2	.372
ages 70+	121	0.612, 38.241	1.39	2	.499
adults (<30+)	1135	0.958, 24.073	4.15	2	.126

Table A9

Details of fitted distributions

Busselton 1975, females [Table 5.14]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages < 30	154	2.261, 0.906	3.03	2	.220
ages 30-39	129	2.276, 0.975	3.00	2	.223
ages 40-49	147	2.325, 0.970	2.82	2	.244
ages 50-59	148	2.631, 0.771	0.08	1	.778
ages 60-69	109	2.779, 0.855	1.90	3	.594
ages 70+	53	2.155, 1.320	2.87	2	.238
adults (<30+)	740	2.443, 0.917	3.22	3	.360
two parameter lognormal with covariance on t = age in years					
adults (<30+)	740	$\mu = 0.9732 + .05683t - .00050t^2$			
		$\sigma = 1.7720 - .03969t + .00042t^2$	20.83	12	.053
t=25		2.082, 1.042	3.90	2	.142
t=35		2.350, 0.896	3.28	2	.194
t=45		2.519, 0.834	4.58	2	.101
t=55		2.588, 0.857	0.74	1	.390
t=65		2.557, 0.963	4.74	3	.192
t=75		2.426, 1.152	3.59	2	.166
two parameter lognormal, censored below 40 g/day					
adults (<30+)	740	2.533, 0.862	2.87	2	.238
three parameter lognormal (μ,σ,τ)					
ages < 30	154	2.927, 0.661, -11.988	2.84	1	.092
ages 30-39	129	0.426, 1.833, 14.160	1.73	1	.188
ages 40-49	147	2.127, 1.055, 2.605	2.80	1	.094
ages 60-69	109	3.920, 0.422, -36.396	0.60	2	.742
ages 70+	53	0.952, 1.978, 10.972	2.56	1	.110
adults (<30+)	740	2.572, 0.863, -2.045	3.16	2	.206
gamma (α,β)					
ages < 30	154	0.629, 19.865	3.17	3	.366
ages 30-39	129	0.549, 24.919	5.40	3	.145
ages 40-49	147	0.587, 24.516	4.07	3	.254
ages 50-59	148	1.135, 15.184	0.07	1	.791
ages 60-69	109	0.992, 21.782	0.71	3	.871
ages 70+	53	0.348, 50.226	3.82	3	.282
adults (<30+)	740	0.698, 22.457	6.94	3	.074
gamma, censored below 40 g/day					
adults (<30+)	740	0.514, 27.255	5.43	2	.066

Table A10

Details of fitted distributions

AWU members [Table 5.15]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males	2662	2.370, 1.008	3.78	1	.052
females	5074	1.680, 0.864	28.59	1	<.001
gamma (α,β)					
males	2662	0.801, 21.492	49.15	1	<.001
females	5074	0.604, 11.534	59.47	1	<.001

Table A11

Details of fitted distributions

Medicheck screenings [Table 5.16]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males 1975	3653	2.822, 0.887	24.32	1	<.001
males 1976	5956	2.766, 0.907	8.19	1	.004
females 1975	1816	2.348, 0.742	6.73	1	.010
females 1976	3446	2.248, 0.736	7.90	1	.005
gamma (α,β)					
males 1975	3653	1.258, 19.500	150.53	1	<.001
males 1976	5956	1.183, 19.876	144.65	1	<.001
females 1975	1816	1.328, 10.422	30.23	1	<.001
females 1976	3446	1.262, 9.777	38.39	1	<.001

Table A12

Details of fitted distributions

Geelong beach survey [Table 5.17]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
male youths (10-19)	383	3.127, 0.979	12.48	1	<.001
males 15-19	355	3.180, 0.972	13.46	1	<.001
males 20-24	105	3.540, 0.926	23.87	1	<.001
males 25+	55	3.184, 1.122	1.19	1	.274
male adults (20+)	160	3.430, 0.990	22.68	1	<.001
female youths (10-19)	284	1.819, 1.207	0.14	1	.708
females 15-19	256	1.690, 1.298	0.03	1	.871
gamma (α,β)					
male youths (10-19)	383	0.833, 39.714	6.32	1	.012
males 15-19	355	0.871, 39.825	7.11	1	.008
males 20-24	105	1.169, 40.601	18.20	1	<.001
males 25+	55	0.707, 55.006	0.51	1	.475
male adults (20+)	160	0.988, 45.147	16.28	1	<.001
female youths (10-19)	284	0.235, 40.535	0.67	1	.413
females 15-19	256	0.202, 46.211	0.35	1	.554

Table A13

Details of fitted distributions
North-west Melbourne [Table 5.18]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
male youths (15-19)	91	1.394, 1.636	2.17	1	.140
males 20-29	208	1.936, 1.677	14.61	1	<.001
males 30-39	172	2.304, 1.508	11.55	1	<.001
males 40-49	155	2.585, 1.865	20.13	1	<.001
males 50-59	117	2.358, 1.460	5.96	1	.015
males 60-69	78	2.312, 1.676	8.24	1	.004
males 70+	32	1.833, 1.577	1.89	1	.169
male adults (20+)	762	2.244, 1.659	58.36	1	<.001
females 20-29	200	0.151, 1.838	2.71	1	.100
female adults (20+)	683	0.812, 1.477	12.09	1	<.001
two parameter lognormal with covariance on t = age in years					
male adults (20+)	762	$\mu = 0.1935 + .0943t - .00096t^2$			
		$\sigma = 1.5621 + .0044t - .00005t^2$	68.58	6	<.001
t=25		1.949, 1.637	14.68	1	<.001
t=35		2.313, 1.649	12.61	1	<.001
t=45		2.485, 1.649	22.87	1	<.001
t=55		2.464, 1.639	8.05	1	.005
t=65		2.250, 1.618	8.48	1	.004
t=75		1.843, 1.586	1.90	1	.169
gamma (α,β)					
male youths (15-19)	91	0.253, 47.985	0.47	1	.493
males 20-29	208	0.331, 61.237	5.87	1	.015
males 30-39	172	0.446, 53.763	4.04	1	.044
males 40-49	155	0.392, 100.000	10.62	1	.001
males 50-59	117	0.475, 50.839	1.64	1	.200
males 60-69	78	0.396, 69.541	3.84	1	.050
males 70+	32	0.335, 50.454	0.70	1	.403
male adults (20+)	762	0.387, 66.578	23.22	1	<.001
females 20-29	200	0.109, 43.403	0.75	1	.387
female adults (20+)	683	0.185, 29.525	4.27	1	.039

Table A14

Details of fitted distributions

Geelong school survey, males [Table 5.19]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
age 14	307	1.561, 1.066	10.53	3	.015
age 15	496	1.811, 1.151	12.78	3	.005
age 16	371	2.086, 0.973	2.26	3	.521
age 17	182	2.152, 0.919	6.38	3	.094
age 18	52	2.464, 0.704	0.77	2	.681
youths (all)	1424	1.921, 1.036	6.56	3	.087
two parameter lognormal, censored below 15 g/day					
youths (all)	1424	1.848, 1.092	4.58	2	.101
two parameter lognormal with covariance on t = age in years					
youths (all)	1424	$\mu = -1.6618 + .24310t - .00083t^2$			
		$\sigma = -5.8417 + .95490t - .03284t^2$	36.99	14	<.001
t=14		1.580, 1.090	10.82	3	.013
t=15		1.799, 1.092	14.27	3	.003
t=16		2.016, 1.029	3.68	3	.298
t=17		2.232, 0.900	7.42	3	.060
t=18		2.446, 0.706	0.79	2	.850
three parameter lognormal (μ,σ,τ)					
youths (all)	1424	1.663, 1.171, 1.734	4.44	2	.109
gamma (α,β)					
age 14	307	0.501, 15.477	22.06	3	<.001
age 15	496	0.326, 55.127	58.31	3	<.001
age 16	371	0.806, 15.352	16.99	3	<.001
age 17	182	0.897, 14.288	17.86	3	<.001
age 18	52	1.831, 8.106	2.91	2	.233
youths (all)	1424	0.660, 16.711	50.48	3	<.001
gamma, censored below 15 g/day					
youths (all)	1424	0.417, 23.730	13.08	2	.001

Table A15

Details of fitted distributions

Geelong school survey, females [Table 5.19]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
age 14	233	1.057, 1.152	2.61	2	.271
age 15	386	1.385, 0.989	8.69	2	.013
age 16	292	1.666, 0.777	7.76	2	.021
age 17	178	1.730, 0.829	0.01	1	.923
youths (all)	1015	1.495, 0.907	10.53	3	.015
two parameter lognormal, censored below 15 g/day					
youths (all)	1127	1.151, 1.106	4.13	2	.127
two parameter lognormal with covariance on t = age in years					
ages 14-17	1103	$\mu = -20.6087 + 2.6240t - .07704t^2$			
		$\sigma = 18.5706 - 2.1615t + .06572t^2$	19.89	7	.006
t=14		1.027, 1.191	2.69	2	.261
t=15		1.416, 0.935	9.17	2	.010
t=16		1.652, 0.811	8.01	2	.018
t=17		1.734, 0.818	0.02	1	.876
three parameter lognormal (μ,σ,τ)					
youths (all)	1127	0.492, 1.343, 4.162	3.54	2	.171
gamma (α,β)					
age 14	233	0.320, 14.793	7.78	3	.051
age 15	386	0.507, 11.384	22.71	2	<.001
age 16	292	1.103, 6.619	18.34	2	<.001
youths (all)	1127	0.629, 9.583	40.00	3	<.001

Table A16

Details of fitted distributions

Newcastle alcohol abusers [Table 5.20]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
all	100	5.344, 0.431	56.95	7	<.001
two parameter lognormal, censored below ζ g/day					
all	100	5.209, 0.594 (ζ = 180)	22.07	6	.001
all	100	5.198, 0.604 (ζ = 240)	22.04	5	<.001
two parameter lognormal, censored below 180 g/day and above 360 g/day					
all	100	5.229, 0.484	2.49	2	.288
three parameter lognormal, censored below ζ g/day					
all	100	4.642, 0.842, 80.739 (ζ = 180)	20.87	5	<.001
all	100	3.475, 1.354, 181.376 (ζ = 240)	18.35	4	.001
gamma (α,β)					
all	100	0.728, 49.383	77.72	10	<.001
gamma, censored below ζ g/day					
all	100	2.168, 97.752 (ζ = 180)	25.93	8	.001
all	100	1.788, 113.636 (ζ = 240)	24.93	7	<.001

Table A17

Details of fitted distributions

ABS survey, males [Table 5.21]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages 18–24	1064	2.890, 1.051	51.71	6	<.001
ages 25–44	2618	2.934, 1.000	109.44	12	<.001
ages 45–64	1753	2.922, 1.044	86.18	8	<.001
ages 65+	464	2.379, 1.051	12.38	6	.054
adults (18+)	5897	2.879, 1.029	235.56	12	<.001
two parameter lognormal, truncated below 20 g/day					
ages 18–24	532	3.369, 0.727	15.56	4	.004
ages 25–44	1345	3.312, 0.760	25.95	10	.004
ages 45–64	901	3.427, 0.708	11.01	6	.088
ages 65+	139	3.291, 0.617	3.59	4	.464
adults (18+)	2915	3.329, 0.752	29.65	10	.001
two parameter lognormal, truncated below ζ g/day					
adults (18+)	1969	3.549, 0.666 ($\zeta = 30$)	18.91	9	.026
adults (18+)	1424	3.226, 0.768 ($\zeta = 40$)	11.57	8	.172
adults (18+)	983	2.733, 0.856 ($\zeta = 50$)	9.02	7	.251
adults (18+)	670	2.516, 0.930 ($\zeta = 60$)	8.90	6	.179
adults (18+)	462	3.043, 0.824 ($\zeta = 70$)	8.44	5	.134
adults (18+)	326	3.906, 0.621 ($\zeta = 80$)	5.72	4	.221
two parameter lognormal, censored below 20 g/day					
ages 18–24	1064	3.007, 0.908	20.67	5	.001
ages 25–44	2618	3.038, 0.883	35.51	11	<.001
ages 45–64	1753	3.043, 0.900	24.44	7	.001
ages 65+	464	2.515, 0.934	7.83	5	.166
adults (18+)	5897	2.996, 0.901	58.75	11	<.001
two parameter lognormal, censored below ζ g/day					
adults (18+)	5897	3.039, 0.864 ($\zeta = 30$)	46.75	10	<.001
adults (18+)	5897	3.130, 0.797 ($\zeta = 40$)	11.79	9	.225
adults (18+)	5897	3.157, 0.780 ($\zeta = 50$)	10.08	8	.259
adults (18+)	5897	3.141, 0.789 ($\zeta = 60$)	9.73	7	.205
adults (18+)	5897	3.100, 0.811 ($\zeta = 70$)	8.44	6	.208
adults (18+)	5897	3.045, 0.838 ($\zeta = 80$)	7.13	5	.211

Table A17
(continued)

Details of fitted distributions

ABS survey, males [Table 5.21]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal censored below 50 g/day with covariance on *t*					
adults (18+)	5897	$\mu = 2.4900 + .0325t - .00034t^2$			
		$\sigma = 0.9606 - .0054t + .00001t^2$	29.08	16	.023
t=21		3.024, 0.852	7.94	2	.019
t=35		3.213, 0.787	13.05	8	.110
t=55		3.253, 0.704	6.66	4	.155
t=75		3.022, 0.632	1.43	2	.488
three parameter lognormal (μ,σ,τ)					
ages 18-24	1064	3.631, 0.604, -17.883	15.52	5	.008
ages 25-44	2618	3.520, 0.660, -13.282	24.73	11	.010
ages 45-64	1753	3.702, 0.587, -19.810	10.35	7	.170
ages 65+	464	3.203, 0.637, -13.882	4.44	5	.487
adults (18+)	5897	3.519, 0.661, -14.321	30.02	11	.002
gamma (α,β)					
ages 18-24	1064	0.982, 28.506	17.90	6	.007
ages 25-44	2618	1.008, 28.944	34.93	12	<.001
ages 45-64	1753	0.980, 29.621	16.22	8	.039
ages 65+	464	0.777, 22.262	3.06	6	.801
adults (18+)	5897	0.941, 30.048	61.69	12	<.001

Table A18

Details of fitted distributions

ABS survey, females [Table 5.21]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages 18-24	757	1.824, 1.001	9.66	4	.047
ages 25-44	1765	1.963, 0.942	3.40	4	.493
ages 45-64	1100	2.035, 1.000	18.84	4	.001
ages 65+	345	1.502, 1.173	10.50	3	.015
adults (18+)	3966	1.925, 0.989	9.33	4	.053
two parameter lognormal, truncated below 10 g/day					
ages 18-24	238	2.439, 0.769	7.72	3	.052
ages 25-44	635	1.803, 0.994	3.22	3	.359
ages 45-64	428	2.751, 0.683	6.92	3	.074
ages 65+	85	2.359, 0.867	9.78	2	.008
adults (18+)	1386	2.371, 0.820	3.24	3	.356
two parameter lognormal, censored below 30 g/day					
ages 18-24	767	2.047, 0.860	7.21	2	.027
ages 25-44	1765	2.018, 0.918	2.29	2	.318
ages 45-64	1100	2.455, 0.723	4.62	2	.099
ages 65+	345	1.296, 1.250	7.74	1	.005
adults (18+)	3966	2.149, 0.852	1.64	2	.440
two parameter lognormal with covariance on t = age in years					
adults (18+)	3966	$\mu = 1.0737 + .0453t - .00052t^2$			
		$\sigma = 1.1930 - .0128t + .00017t^2$	50.65	15	<.001
t=21		1.795, 0.998	10.16	4	.038
t=35		2.019, 0.951	8.67	4	.070
t=55		1.982, 0.997	21.20	4	<.001
t=75		1.527, 1.178	10.62	3	.014
three parameter lognormal (μ,σ,τ)					
ages 18-24	757	2.384, 0.759, -5.648	7.93	3	.048
ages 25-44	1765	1.864, 0.988, 0.805	3.24	3	.356
ages 45-64	1100	3.065, 0.557, -15.064	6.44	3	.092
ages 65+	345	2.090, 0.912, -5.118	9.90	2	.007
adults (18+)	3966	2.341, 0.802, -4.169	3.16	3	.368
gamma (α,β)					
ages 18-24	757	0.634, 14.545	10.37	4	.035
ages 25-44	1765	0.749, 13.650	20.92	4	<.001
ages 45-64	1100	0.724, 15.931	6.44	4	.169
ages 65+	345	0.408, 18.914	11.41	4	.022
adults (18+)	3966	0.684, 14.914	15.52	4	.004

Table A19

Details of fitted distributions

Adelaide University students [Table 5.23]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males	108	1.987, 0.642	1.40	2	.496
two parameter lognormal, censored below 20 g/day					
males	108	1.278, 1.021	0.003	1	.959
three parameter lognormal (μ,σ,τ)					
males	108	0.034, 1.462, 7.906	0.004	1	.948
gamma (α,β)					
males	108	1.530, 5.469	3.09	2	.213

Table A20

Details of fitted distributions

Busselton 1978, males [Table 5.24]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages <30	249	3.071, 0.794	3.42	3	.331
ages 30-39	237	2.931, 0.797	2.40	3	.493
ages 40-49	205	2.896, 0.944	7.17	3	.067
ages 50-59	209	3.118, 0.884	14.60	3	.002
ages 60-69	202	2.915, 0.982	8.36	3	.040
ages 70+	158	2.790, 0.835	0.47	3	.925
adults (<30+)	1260	2.965, 0.876	8.40	3	.039
two parameter lognormal, truncated below 20 g/day					
ages < 30	135	2.476, 1.013	2.57	2	.276
ages 30-39	110	3.309, 0.629	1.41	2	.495
ages 40-49	94	2.815, 0.975	7.16	2	.028
ages 50-59	112	3.763, 0.486	0.54	2	.762
ages 60-69	92	3.705, 0.549	0.07	2	.967
ages 70+	64	1.737, 1.147	0.06	2	.969
adults (<30+)	607	3.356, 0.696	2.67	2	.263
two parameter lognormal, censored below 40 g/day					
ages < 30	249	2.986, 0.863	2.82	2	.244
ages 30-39	237	3.064, 0.698	1.20	2	.549
ages 40-49	205	2.852, 0.978	7.08	2	.029
ages 50-59	209	3.399, 0.625	0.43	2	.807
ages 60-69	202	3.217, 0.730	0.65	2	.721
ages 70+	158	2.658, 0.925	0.13	2	.936
adults (<30+)	1260	3.083, 0.781	2.47	2	.291
two parameter lognormal with covariance on t **= age in years**					
adults (<30+)	1260	$\mu = 2.9326 + .0057t - .00009t^2$ $\sigma = 0.4307 + .0179t - .00016t^2$	49.60	18	<.001
t=25		3.015, 0.778	4.82	3	.159
t=35		3.015, 0.861	6.69	3	.083
t=45		2.996, 0.912	8.67	3	.034
t=55		2.959, 0.931	18.89	3	<.001
t=65		2.902, 0.918	9.35	3	.025
t=75		2.827, 0.873	1.18	3	.758

Table A20
(continued)

Details of fitted distributions

Busselton 1978, males [Table 5.24]

Sample	n	parameter estimates	χ^2	df	prob
three parameter lognormal (μ,σ,τ)					
ages < 30	249	2.689, 0.995, 6.757	2.55	2	.279
ages 30-39	237	3.431, 0.584, -12.554	1.55	2	.460
ages 40-49	205	2.851, 0.967, 0.851	7.16	2	.028
ages 70+	158	2.303, 1.067, 7.072	0.06	2	.968
adults (<30+)	1260	3.486, 0.635, -13.622	2.82	2	.245
gamma (α,β)					
ages < 30	249	1.307, 21.730	9.40	3	.024
ages 30-39	237	1.226, 19.928	2.16	3	.540
ages 40-49	205	0.880, 29.949	8.76	3	.033
ages 50-59	209	1.147, 27.181	6.21	3	.102
ages 60-69	202	0.858, 31.867	2.44	3	.486
ages 70+	158	1.014, 21.363	3.18	3	.365
adults (<30+)	1260	1.052, 25.426	4.46	3	.216

Table A21

Details of fitted distributions

Busselton 1978, females [Table 5.24]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
ages < 30	201	1.927, 0.964	0.60	1	.438
ages 30-39	203	1.845, 1.017	5.67	1	.017
ages 40-49	168	2.340, 0.852	0.14	1	.708
ages 60-69	168	2.335, 1.017	3.38	3	.337
ages 70+	80	2.138, 0.825	0.46	1	.496
adults (<30+)	1001	2.272, 0.853	4.56	3	.207
two parameter lognormal with covariance on t = age in years					
adults (<30+)	820	$\mu = .7298 + .0564t - .000505t^2$			
except 50-59		$\sigma = .9150 + .0023t - .000029t^2$	17.86	7	.013
t=25		1.825, 0.955	1.76	1	.185
t=35		2.086, 0.961	8.50	1	.004
t=45		2.247, 0.961	0.74	1	.389
t=65		2.264, 0.943	5.81	3	.121
t=75		2.121, 0.925	1.05	1	.305
two parameter lognormal, censored below 40 g/day					
adults (<30+)	1001	2.623, 0.666	0.40	2	.818
three parameter lognormal (μ,σ,τ)					
adults (<30+)	1001	3.564, 0.436, -31.198	1.11	2	.573
gamma (α,β)					
ages < 30	201	0.456, 19.301	1.42	1	.233
ages 30-39	203	0.394, 21.372	4.74	2	.094
ages 40-49	168	0.774, 16.932	0.01	1	.920
ages 60-69	168	0.561, 27.137	2.86	3	.414
ages 70+	80	0.673, 14.734	0.88	1	.348
adults (<30+)	1001	0.698, 17.232	1.48	3	.687

Table A22

Details of fitted distributions

RAAF recruits [Table 5.25]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
incoming 17–20	260	2.765, 0.826	1.67	2	.435
incoming 21–25	107	3.173, 0.763	1.70	2	.428
incoming 26+	70	2.680, 1.123	1.22	2	.543
incoming adults (all)	444	2.867, 0.859	2.73	2	.255
outgoing 17–20	286	3.118, 0.802	0.45	2	.798
outgoing 21–25	121	3.211, 0.736	1.59	2	.453
outgoing 26+	73	3.247, 0.863	1.52	2	.469
outgoing adults (all)	491	3.158, 0.800	0.86	2	.652
two parameter lognormal with covariance on t **= age in years**					
incoming adults (all)	437	$\mu = 2.9138 - .000085t$			
		$\sigma = 0.2435 + .027916t$	12.16	6	.059
incoming t=18.5		2.912, 0.760	3.92	2	.159
incoming t=23		2.912, 0.886	6.01	2	.050
incoming t=29.5		2.911, 1.011	2.22	2	.330
outgoing adults (all)	480	$\mu = 2.8729 + .01388t$			
		$\sigma = 0.6465 + .00669t$	4.32	6	.634
outgoing t=18.5		3.130, 0.770	0.70	2	.706
outgoing t=23		3.192, 0.800	2.04	2	.361
outgoing t=29.5		3.255, 0.830	1.58	2	.454
two parameter lognormal, censored below 50 g/day					
incoming 17–20	260	3.020, 0.672	0.38	1	.538
incoming 21–25	107	3.180, 0.757	1.70	1	.193
incoming 26+	70	3.059, 0.850	0.13	1	.722
incoming adults (all)	444	3.049, 0.738	1.17	1	.278
outgoing 17–20	286	3.017, 0.877	0.00	1	.979
outgoing 21–25	121	3.307, 0.662	1.26	1	.262
outgoing 26+	73	3.427, 0.703	0.52	1	.471
outgoing adults (all)	491	3.155, 0.802	0.85	1	.355

Table A22
(continued)

Details of fitted distributions

RAAF recruits [Table 5.25]

Sample	n	parameter estimates	χ^2	df	prob
three parameter lognormal (μ,σ,τ)					
incoming 21-25	107	2.918, 0.881, 5.972	1.64	1	.200
incoming adults (all)	444	3.661, 0.555, -25.050	1.86	1	.173
outgoing 17-20	286	2.633, 1.036, 10.024	0.03	1	.870
outgoing 21-25	121	3.520, 0.608, -9.603	1.50	1	.221
outgoing 26+	73	4.292, 0.426, -49.200	0.96	1	.327
outgoing adults (all)	491	3.033, 0.858, 3.068	0.80	1	.371
gamma (α,β)					
incoming 17-20	260	0.844, 23.288	0.91	2	.634
incoming 21-25	107	1.285, 23.180	2.48	2	.289
incoming 26+	70	0.515, 45.413	0.48	2	.787
incoming adults (all)	444	0.856, 26.560	2.07	2	.355
outgoing 17-20	286	1.132, 25.491	2.57	2	.277
outgoing 21-25	121	1.405, 21.716	1.75	2	.417
outgoing 26+	73	1.100, 31.201	1.07	2	.586
outgoing adults (all)	491	1.171, 25.641	3.31	2	.191

Table A23

Details of fitted distributions

Queensland human service students [Table 5.26]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males	685	2.488, 1.053	1.04	2	.595
females	583	1.951, 0.896	0.58	2	.746
two parameter lognormal, censored below 40 g/day					
males	685	2.530, 1.024	0.95	1	.329
females	583	2.120, 0.817	0.43	1	.511
three parameter lognormal (μ,σ,τ)					
males	685	2.755, 0.926, -4.616	0.81	1	.369
females	583	2.355, 0.760, -5.561	0.51	1	.476
gamma (α,β)					
males	685	0.596, 31.456	2.27	2	.321
females	583	0.492, 16.915	1.01	2	.604
gamma, censored below 40 g/day					
males	685	0.439, 39.185	0.17	1	.681
females	583	15.326, 1.777	56.22	1	<.001

Table A24

Details of fitted distributions

Townsville residents [Table 5.27]

Sample	n	parameter estimates	χ^2	df	prob
two parameter lognormal (μ,σ)					
males 18-24	35	3.855, 0.542	0.67	1	.413
males 25-44	44	3.812, 0.767	4.45	1	.035
males 45-64	19	3.392, 0.911	0.27	1	.601
male adults (18+)	110	3.761, 0.680	1.24	1	.266
gamma (α,β)					
males 18-24	35	3.029, 17.778	1.39	1	.238
males 25-44	44	1.587, 36.049	2.85	1	.091
males 45-64	19	0.914, 44.863	0.09	1	.764
male adults (18+)	110	1.894, 27.465	0.23	1	.632
female adults (18+)	76	0.559, 29.735	3.57	1	.059

PART II

Chapter 7

Inference on linear functions of class probabilities.

7.1 Introduction

In Part I of this thesis we considered the problem of choosing a specification to describe aspects of the distribution of individual alcohol consumption. We demonstrated that censoring the lower tail of a lognormal distribution gave a better fit in the upper tail of the distribution, as did adding a lower threshold parameter to the specification.

In this second part we will consider these and related inferential problems in a more mathematical fashion. At the base of the matter lie the principles of relevance and noncoherence, as given by Wilkinson (1977). For the relevance principle requires that any inferential statements be made on the basis of all the relevant information in the sample, excluding all irrelevant or spurious information. The determination of which information is relevant can only be made by a precise formulation of the questions which the inference is designed to answer. The noncoherence principle implies that the inference may change radically in response to slight changes in the question (James, 1977).

Related to these principles are two aspects of the statistical analysis of data: value and validity. By value we refer to the information in the data and its relevance as mentioned above. Validity includes the goodness-of-fit of the models by which we interpret the data.

For frequency distributions, a linear functional of the frequencies or probabilities can highlight either value or validity. Functionals whose domain is restricted to that of the probability vector, that is, to the estimation locus, will be functions of the parameters or their estimates, and hence express aspects of inference assuming the parametric specification. For example, in our studies of the distribution of alcohol consumption three important linear functionals highlighting three different values are

i. the mean alcohol consumption.

ii. the proportion of heavy drinkers, for example, above 60 g/day.

iii. an index of excess consumption over and above, say, 60 g/day.

On the question of validity of a proposed specification, functionals of the deviations $(\mathbf{f}-\hat{\mathbf{p}})$ of the relative frequency vector from the estimated probability vector will express aspects of the goodness-of-fit of the data to the specification. Certain functionals determine the components of a χ^2 goodness-of-fit test. If a component is significant, the specification must be amended to incorporate the significant effect. However we shall be concerned with the situation in which, although a component of χ^2 may not be significant, it is imprudent to trust that there is no real effect.

Our approach is to decompose a linear functional to show that a non-parametric estimator of a particular contrast or value is partitioned into the parametric estimator plus a second component whose expected value is zero, the variances of these two components being additive. In gaining the advantage of the smaller variance of the parametric estimator, we are depending on the validity of the specification to assume that the second component has zero expectation. If we have some doubt as to the validity of a particular aspect of the parametric specification, for example the symmetry of the log-normal distribution on the log scale, we may modify the specification and transfer a further component from the second component to the parametric estimator, and be confident that the expectation of the new, reduced second component is zero. We would then have greater confidence in the validity of the modified specification.

We will show that, in the case of estimation of the distribution of alcohol consumption by the two parameter lognormal, modifying the specification by the addition of the third parameter, or altering the the fitting procedure by censoring the lower class frequencies, may ensure validity with respect to certain values. An example will be presented where the two alternatives have almost identical effect.

Questions of validity with respect to different values may come into conflict; there may be no single fitted distribution which has optimal validity in respect of disparate values. This is an example of the noncoherence principle.

The formulation of these important qualitative inferential ideas in a precise quantitative manner using linear functionals of relative frequencies and probabilities requires a very careful algebraic treatment. For example, suppose we estimate the proportion of drinkers consuming more than 60 g alcohol per day. Since the probabilities and relative frequencies add to unity, an estimate of this proportion implies equivalently an estimate of the proportion of consumers of *less* than 60 g per day, and the variance of the estimate of the complementary proportion will be the same as for the estimate of the proportion.

This point is made to show that there is a certain redundancy in the problem and this must enter into its algebraic formulation. We shall develop the mathematics of linear functionals in a later section, but firstly explore some general considerations in the choice of a suitable specification for the distribution of alcohol consumption, and introduce some linear functionals relevant to this study.

7.2 The choice of a specification for the distribution of alcohol consumption

We can distinguish several broad motivations for fitting a mathematical distribution to a set of data:

i. To summarise and describe a situation, when we wish to smooth the class frequencies. The summary may be for its own sake, or to examine an hypothesis involving the distributional form as a consequence, such as the hypothesis that alcoholics are not essentially different from other drinkers, or more precisely, that the distribution of consumption is not bimodal.

ii. To enable estimation of population characteristics not readily obtainable directly, for example, for small and moderate sized samples, the extrapolation to relative frequencies in the extreme upper tail from a fit based primarily on the middle and low upper-tail frequencies. While this procedure is subject to the usual uncertainties and doubts of extrapolation, if a specification has been established from large data sets with reasonable absolute frequencies in the upper tail, then in the absence of information to the contrary, the best inference one can make for small samples is to assume that they will be similar and use estimates based on the fitted specification.

iii. Various reasons such as error estimates, tests between and within samples, interpolation and graduation of frequencies etc.

Researchers in the alcohol field have tried to find a single specification for the distribution of individual consumption which would serve all these needs.

"A Statistical Study of the Distribution of Alcohol Consumption and Consequent Inferential Problems" by J.B.F. Field

Erratum

page 186, line 17: "left" should read "right"

However the choice of a specification or the method of fitting should depend upon which aspects of the data are considered important for inference. A specification which is satisfactory for one purpose may be quite unsuited to another. Even a "good fit" is not necessarily a sound basis for inference, in that the test of fit does not establish the distribution, it merely assesses the evidence *against* this specification. It is quite possible for a particular distribution to be in close agreement with a data set for the central 80–90% but be in disagreement with the population in the tails, where less data is available to assess the fit. This situation may be acceptable if we wish to make inferences about, say, the mean of the distribution, but inferences about the probabilities in the tails of the distribution are much more sensitive to the specification.

Since alcohol consumption is inherently non-negative, and since the overwhelming majority consume miniscule amounts of alcohol in relation to the amounts consumed by the still appreciable minority of problem drinkers, it is clear *à priori* that the distribution of alcohol consumption will be skewed to the left. The most common way of accommodating this skewness is to postulate that the logarithm of consumption has a symmetric distribution, say normal.

This postulate is usually, in gross terms, highly effective. We do not wish to imply that the lognormal distribution is correct, but merely that the log transformation results in a good visual description of the distribution, and that the residual asymmetry is so slight, and indeed the accuracy of recording so coarse, that any formal test of fit would require a very large sample to register significant evidence of departure.

More importantly, however, any fitted symmetric distribution to the transformed data will effect a compromise between the upper and lower tails of the distribution. Unfortunately in the case of alcohol studies the interest in the distribution is not so symmetric and non-committal, but rather concentrated on the upper tail of the distribution. This leads to something of a dilemma, because inferences about heavy drinkers can be influenced substantially by inconsequential variations in the habits of the light drinkers; that is, information about light drinkers is being spuriously used to make inferences about heavy drinkers.

To demonstrate this point consider some real data. Table 7.1 (a subset of Table 5.24) shows a frequency distribution for the alcohol consumption for 1001 females of all ages in the 1978 Busselton, W.A., survey (Cullen *et al*, 1980). For convenience, we shall refer to this data as the "Busselton data" throughout this chapter.

Table 7.1

Effect of a small adjustment to the lower tail of the two parameter lognormal distribution. Busselton, W.A., females, 1978

class	Original data			Adjusted data	
int.	nf	f	$\hat{p}$	nf	$\hat{p}$
1-20	804	.8032	.8020	814	.8119
21-40	142	.1419	.1497	132	.1407
41-60	42	.0420	.0320	42	.0308
61-80	9	.0090	.0096	9	.0096
81-100	2	.0020	.0036	2	.0037
>100	2	.0020	.0031	2	.0034

Suppose we "adjust" the frequencies by moving 10 people (about 1% of the sample) from the 21-40 class interval to the 1-20 interval. We fit two parameter lognormal distributions to both the original and adjusted data. The parameter estimates change from (μ, σ) = (2.2717, 0.8529) to (2.2146, 0.8826), and the χ^2 goodness-of-fit test statistic on 3 degrees of freedom changes from 4.56 (P = 0.207) to 5.92 (P = 0.115). Table 7.1 shows the frequencies and fitted values for the two cases. The small adjustment to the lower tail has changed the combined probabilities for the two upper classes from 0.0067 to 0.0071, a 6% change. Similarly, an adjustment of 2% in the lower tail produces a change of 10% in the upper tail. Thus the area of most importance to us, the upper tail, has been substantially affected by a small change in which we have little interest, in the lower tail.

To compensate for this anomalous situation, we look for a mode of inference which is less sensitive to perturbations in the lower tail when the inferential emphasis is on the upper tail.

There are good historical precedents for choosing a particular parametric specification in order to obtain prescribed estimators as the maximum likelihood ones. Gauss (1809) assumed that "when any number of equally good direct observations M, M', M'', ... of an unknown magnitude x are given, the most probable value is their arithmetic mean" (Whittacker and Robinson, 1932). Using this postulate he then deduced that the observations must be Normally distributed about the true value. Von Mises (1918) asked "For what distribution on the unit circle is the unit vector $\hat{\mu} = (\cos\theta_0 \ \sin\theta_0)$ a maximum likelihood estimator of a direction θ_0 of clustering or concentration?" (Bingham, 1980). And Fisher (1953) in his paper on the distribution of dispersion on the sphere, similarly chose his specification in order to obtain as the maximum likelihood estimator of location the three dimensional analogue of Von Mises' $\hat{\mu}$.

Various solutions are worthy of consideration in the present case. The simplest is possibly just to truncate the distribution at some arbitrary point, and consider only drinkers whose consumption exceeds (say) 60 g alcohol per day. This procedure suffers from two drawbacks: the choice of the truncation point is arbitrary, and more seriously, truncating any more than the consumptions of the very light drinkers throws away most of the data, as reference to most of the tables of consumption data in Chapter 6 will show.

A less severe alternative is to censor the data rather than truncate it. That is, we assume we know only the proportion of the distribution lying below the point of censorship (60 g alcohol per day, say) and have no detailed knowledge of the consumption values for this portion of the data. For the typical grouped data that is available from alcohol consumption surveys, this amounts to combining the lower class intervals into one class. This solution suffers from the same problem of arbitrariness as does truncation, but has the great advantage that we are not discarding the information about the light and moderate drinkers entirely.

Another solution of the problem is to introduce a lower threshold parameter, τ, to the specification, and so fit a three parameter lognormal distribution. Heuristically, we would expect the threshold parameter to be mainly determined by the smaller observations and thus use more of the information contained in these values than in larger observations. Since the information in the smaller observations is largely "used up" in the estimation of τ, it will be removed, to a considerable extent, from the estimation of the two remaining parameters, μ and σ, which consequently will depend more heavily on the larger observations. These points will be demonstrated quantitatively later when the requisite mathematical machinery has been set up.

A similar situation arises in the context of mining. For about a decade prior to 1960, many gold, uranium and pyrite value distributions in South Africa were estimated using the two parameter lognormal distribution, first introduced* by Sichell (1947). However Krige (1960) showed that there was usually a systematic departure from this model, the departure being in the lower tail. When considered on a log scale the data showed a negative skewness, leading to a positive bias for ore grade estimates. Krige advocated using a threshold parameter to overcome this problem. This three parameter model made significant changes to the lower tail of the distribution with only small adjustments to the upper tail, and it reduced the bias in the mean ore grade estimates. Moreover it was found that the estimated mean ore grade values were not very sensitive to changes in the value of the threshold parameter ranging from close below to well in excess of the optimum value (Krige, 1961; Link and Koch, 1975). The introduction of a third parameter into the specification did not change the overall form of the distribution very much; however we show how it radically changes the fitting of it, leading to more reliable inferences in the middle and upper ranges.

---

* Rasumovsky (1940) had earlier established that ore samples could be represented best by this model. This was unknown to Sichell at the time.

7.3 Linear functionals relevant to alcohol studies

In the introduction to this chapter, we noted some important linear functionals highlighting different values relevant to studies of the distribution of alcohol consumption. In this section we will consider these in relation to the Busselton data, to set the scene for a further mathematical development of linear functionals and the estimators of their values.

a. "Mean consumption". Suppose we wish to estimate the mean consumption of the Busselton data of Table 7.1. The parameter to be estimated can be written as $\mathbf{q}^{*\prime}\mathbf{p}$; $\mathbf{q}^*$ is the linear functional and $\mathbf{q}^{*\prime}\mathbf{p}$ is its value. In this case, $\mathbf{q}^* = \mathbf{q}^*_t$ as given in Table 7.2, with the elements of $\mathbf{q}^*_t$ being the midpoints of the corresponding class intervals. ($n\,\mathbf{q}^{*\prime}_t\mathbf{p}$ is the total consumption; hence the subscript t).

Table 7.2

Some linear functionals relevant to alcohol consumption

class interval	$\mathbf{q}^*_t$	$\mathbf{q}^*_h$	$\mathbf{q}^*_e$
1-20	10	0	0
21-40	30	0	0
41-60	50	0	0
61-80	70	1	10
81-100	90	1	30
>100	110	1	50

The value of the final element q^*_m of $\mathbf{q}^*$ poses some problems, in that the final class interval has only a lower bound, x_m. A satisfactory pragmatic approach is to take q^*_m to be the same distance above

x_m as q^*_{m-1} is below it, i.e.

$$q^*_m = x_m + (x_m - q^*_{m-1})$$

There are two obvious estimators of the value of $\mathbf{q}^{*\prime}\mathbf{p}$: $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ and $\mathbf{q}^{*\prime}\mathbf{f}$ are the parametric and nonparametric estimators respectively. The value of the parametric estimator, $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$, will of course vary with the distribution fitted to the data.

For the Busselton data we have $\mathbf{q}_t^{*\prime}\mathbf{f}$ = 15.415 g alcohol per day, and assuming a two parameter lognormal model (μ = 2.272, σ = 0.853) we have $\mathbf{q}_t^{*\prime}\hat{\mathbf{p}}$ = 15.447. We note that this is in reasonable agreement with $\mathbf{q}_t^{*\prime}\mathbf{f}$, i.e. the difference $\mathbf{q}_t^{*\prime}(\mathbf{f}-\hat{\mathbf{p}})$ is small in relation to $\mathbf{q}_t^{*\prime}\mathbf{f}$ (in fact, only 2% of it).

b. "Number of heavy consumers". Suppose the linear functional consists of unit elements corresponding to the classes designated "heavy drinkers", and zeros elsewhere. Then $n\mathbf{q}^{*\prime}\mathbf{p}$ is the number of heavy consumers, $n\mathbf{q}^{*\prime}\mathbf{f}$ is the number estimated from the data, and $n\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ is the estimate from the fitted distribution.

If we define "heavy consumption" as consumption in excess of 60 g alcohol per day, then for the Busselton data, $\mathbf{q}^*$ is given by $\mathbf{q}_h^*$ of Table 7.2; the estimate from the data is $n\mathbf{q}_h^{*\prime}\mathbf{f}$ = 13, i.e. from the sample we estimate there are 13 women who would be classed as "heavy drinkers" according to our definition. The estimate of this number from the fitted two parameter lognormal distribution is $n\mathbf{q}_h^{*\prime}\hat{\mathbf{p}}$ = 16.34 women. Thus $n\mathbf{q}_h^{*\prime}(\mathbf{f}-\hat{\mathbf{p}})$ = −3.34, or 26% of $n\mathbf{q}_h^{*\prime}\mathbf{f}$.

By adjusting the position of the unit elements in the vector in the obvious way, this functional could estimate the numer of light or medium drinkers in a population. But it is the heavy drinkers in whom we are interested in this study.

c. "<u>Excess consumption</u>". If we put

$$q_i^* = \begin{cases} \text{class midpt} - x_\ell & \text{if consumption} > x_\ell \\ 0 & \text{otherwise} \end{cases}$$

then $\mathbf{q}^{*\prime}\mathbf{p}$ represents consumption in excess of x_ℓ g alcohol/day, with corresponding interpretations for $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ and $\mathbf{q}^{*\prime}\mathbf{f}$. The vector $\mathbf{q}_e^*$ in Table 7.2 is the linear functional which gives consumption in excess of 60 g/day for the Busselton data. We have $\mathbf{q}_e^{*\prime}\mathbf{f} =$ 0.250 g/day, i.e. the average daily consumption in excess of 60 g is 0.25 g per person, as determined by the data. The fitted two parameter lognormal distribution gives an estimate of 0.359 g alcohol per person per day, giving $\mathbf{q}_e^{*\prime}(\mathbf{f}-\hat{\mathbf{p}}) = -0.109$. This is a 44% difference from $\mathbf{q}_e^{*\prime}\mathbf{f}$, as compared to the 2% difference we noted for the mean consumption.

7.4 Linear algebra for estimation from grouped data – preliminaries

7.4.1 Basic definitions and notation Consider a variate X and a sample of size n, $x_1, x_2, \ldots, x_n$. Suppose the data are grouped into m classes. Then we have a (column) vector of frequencies

$$\mathbf{a} = (a_1\ a_2\ \ldots\ a_m)'$$

$$\sum_{i=1}^{m} a_i = n$$

or

$$\mathbf{1}_m'\mathbf{a} = n$$

where $\mathbf{1}_m$ is a vector of unities. In general we work with the relative frequencies

$$\mathbf{f} = \frac{1}{n}\mathbf{a}\ , \quad \mathbf{1}_m'\mathbf{f} = 1.$$

The expectations of the relative frequencies are the class probabilities

$$E[\mathbf{f}] = \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}) \tag{7.01}$$

determined by the probability distribution of X, with

$$\boldsymbol{\theta} = (\theta_1\ \theta_2\ \ldots\ \theta_k)'$$

being the parameters of the distribution. At times we will need to write the probability vector $\mathbf{p}$ as a diagonal matrix, and write

$$\mathbf{P} = \text{diag}(\ \mathbf{p}\ )$$

Let $\mathbf{X}(\boldsymbol{\theta})$ be the $m \times k$ matrix of derivatives of the probabilities $\mathbf{p}$ with respect to the parameters, that is

$$\mathbf{X} = \mathbf{X}(\boldsymbol{\theta}) = [\frac{\partial \mathbf{p}}{\partial \theta_1}\ \frac{\partial \mathbf{p}}{\partial \theta_2}\ \ldots\ \frac{\partial \mathbf{p}}{\partial \theta_k}]$$

$$= [\mathbf{p}_{\theta_1}\ \mathbf{p}_{\theta_2}\ \ldots\ \mathbf{p}_{\theta_k}]$$

Note that since

$$\sum_{j=1}^{m} p_j = 1$$

identically, it follows that

$$\sum_{j=1}^{m} \frac{\partial p_j}{\partial \theta_a} = \sum_{j=1}^{m} \frac{\partial^2 p_j}{\partial \theta_a \partial \theta_b} = 0 \qquad (7.02)$$

Now the frequencies **a** are assumed to be multinomially distributed, $M(n,\mathbf{p}(\boldsymbol{\theta}))$. Thus the likelihood function is

$$\mathcal{L} = \Pr(\mathbf{a}) = \frac{n!}{a_1! \ldots a_m!} p_1^{a_1} \ldots p_m^{a_m}$$

and the variance-covariance matrix $\boldsymbol{\Sigma}$ is defined by

$$\mathrm{var}(\mathbf{f}) = \frac{1}{n}\boldsymbol{\Sigma} = \frac{1}{n}(\mathbf{P} - \mathbf{pp}') .$$

In terms of the relative frequencies, **f**, the likelihood function is

$$\mathcal{L} = \frac{n!}{(nf_1)! \ldots (nf_m)!} p_1^{nf_1} \ldots p_m^{nf_m}$$

and the loglikelihood function is

$$\log \mathcal{L} = \text{constant} + n \sum_{j=1}^{m} f_j \log p_j$$

or, in an obvious matrix notation,

$$= \text{constant} + n\mathbf{f}' \log \mathbf{p}(\boldsymbol{\theta})$$

Differentiating with respect to θ_a gives the ath score component

$$S_a(\boldsymbol{\theta}) = \frac{\partial \log \mathcal{L}}{\partial \theta_a} = n \sum_{j=1}^{m} \frac{f_j}{p_j} \frac{\partial p_j}{\partial \theta_a}$$

and we can write the vector of k score components, the *score vector*, as

$$\mathbf{S}(\boldsymbol{\theta}) = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}}$$

$$= n\mathbf{X}'\mathbf{P}^{-1}\mathbf{f} . \qquad (7.03)$$

A second differentiation of the loglikelihood function gives

$$-\frac{\partial^2 \log \mathcal{L}}{\partial\theta_a \partial\theta_b} = n \sum_{j=1}^{m} \left\{ \frac{f_j}{p_j^2} \frac{\partial p_j}{\partial\theta_a} \frac{\partial p_j}{\partial\theta_b} - \frac{f_j}{p_j} \frac{\partial^2 p_j}{\partial\theta_a \partial\theta_b} \right\} .$$

Taking the expectation, and using (7.01) and (7.02) gives the *information matrix*

$$\begin{aligned} \mathbf{I}(\boldsymbol{\theta}) &= \mathrm{E}\left[-\frac{\partial^2 \log \mathcal{L}}{\partial\boldsymbol{\theta}^2} \right] \\ &= n\left[\left(\sum_{j=1}^{m} \frac{1}{p_j} \frac{\partial p_j}{\partial\theta_a} \frac{\partial p_j}{\partial\theta_b} \right) \right] \\ &= n\mathbf{X}'\mathbf{P}^{-1}\mathbf{X} \end{aligned} \tag{7.04}$$

Equating the score vector (7.03) to zero, we can write the likelihood equations as

$$\mathbf{X}'\mathbf{P}^{-1}\mathbf{f} = \mathbf{0}_k \tag{7.05}$$

The solution of these equations is the maximum likelihood estimate, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$.

7.4.2 <u>Asymptotic assumption</u> The matrices $\mathbf{P}$, $\boldsymbol{\Sigma}$ and $\mathbf{X}$ are functions of the parameter variable $\boldsymbol{\theta}$. It would be ideal to consider them at the true value $\boldsymbol{\theta}^*$ of the parameter. However, as $\boldsymbol{\theta}^*$ is unknown in our example of the distribution of alcohol consumption, we cannot make the numerical calculations we require later on with unknown $\mathbf{P}$, $\boldsymbol{\Sigma}$ and $\mathbf{X}$.

If we take them at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, they become random. A random geometry is unnecessarily complicated and inferentially inappropriate.

Consequently, we shall make the usual asymptotic assumption that there is a domain of parameter space including $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}$ for which variations of $\mathbf{P}$, $\boldsymbol{\Sigma}$ and $\mathbf{X}$ are negligible. We take $\mathbf{P}$, $\boldsymbol{\Sigma}$ and $\mathbf{X}$ at an initial value $\boldsymbol{\theta}_0$ within this domain, and write $\mathbf{p}_0 = \mathbf{p}(\boldsymbol{\theta}_0)$, $\mathbf{p}^* = \mathbf{p}(\boldsymbol{\theta}^*)$, $\hat{\mathbf{p}} = \mathbf{p}(\hat{\boldsymbol{\theta}})$, with similar notation for $\mathbf{P}$, $\boldsymbol{\Sigma}$ and $\mathbf{X}$.

7.4.3 Maximum likelihood estimation as iterated weighted regression

Although it is well known that maximum likelihood estimation can be effected as iterated weighted regression (e.g. Bliss, 1935; Finney, 1952; Fisher, 1954; Nelder and Wedderburn, 1972; Cox and McCullagh, 1982; Green, 1984), we outline the theory for completeness.

To solve the likelihood equations, we can use the method of scoring for parameters (Fisher, 1935, 1954). That is, we choose an initial estimate $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$, close to $\hat{\boldsymbol{\theta}}$, and setting $r = 0$ in the following equation, calculate a new estimate, $\boldsymbol{\theta}_1$:

$$\boldsymbol{\theta}_{r+1} = \boldsymbol{\theta}_r + \mathbf{I}^{-1}(\boldsymbol{\theta}_r)\,\mathbf{S}(\boldsymbol{\theta}_r) \tag{7.06}$$

The process is continued iteratively until the desired accuracy is reached.

We now show that this iterative scheme is equivalent to iterated weighted regression of $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{f} - \mathbf{p}$ on $\mathbf{X}$, with $\mathbf{P}^{-1}$ as the weight matrix.

Since $\mathbf{X}'\mathbf{P}^{-1}\mathbf{p} = \mathbf{0}_k$, we can write the score vector (7.03) as

$$\mathbf{S}(\boldsymbol{\theta}) = n\,\mathbf{X}'\mathbf{P}^{-1}(\mathbf{f} - \mathbf{p}) \tag{7.07}$$

Then by (7.07) and (7.04), the iterative scheme (7.06) becomes

$$\boldsymbol{\theta}_{r+1} = \boldsymbol{\theta}_r + (\mathbf{X}_r'\mathbf{P}_r^{-1}\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{P}_r^{-1}(\mathbf{f} - \mathbf{p}_r) \tag{7.08}$$

Multiplying throughout by $\mathbf{X}_r'\mathbf{P}_r^{-1}\mathbf{X}_r$ gives

$$(\mathbf{X}_r'\mathbf{P}_r^{-1}\mathbf{X}_r)\,\boldsymbol{\theta}_{r+1} = \mathbf{X}_r'\mathbf{P}_r^{-1}\,(\mathbf{X}_r\boldsymbol{\theta}_r + \mathbf{f} - \mathbf{p}_r) \tag{7.09}$$

Let

$$\mathbf{y}_r = \mathbf{X}_r\boldsymbol{\theta}_r + \mathbf{f} - \mathbf{p}_r \tag{7.10}$$

and then we can write equation (7.09) as

$$(X_r'P_r^{-1}X_r)\,\theta_{r+1} = X_r'P_r^{-1}y_r$$

which are the normal equations for an iterated weighted regression of y_r on X_r with weight matrix P_r^{-1}. By the usual regression theory, the solutions at the rth iteration will be given by

$$\theta_{r+1} = (X_r'P_r^{-1}X_r)^{-1}X_r'P_r^{-1}y_r$$

which, using (7.10), is equation (7.08).

If we use a one-step approximation, then from (7.08) we can write

$$\hat{\theta} \simeq \theta_0 + (X'P^{-1}X)^{-1}X'P^{-1}(f - p_0)$$

where X and P are evaluated at $\theta = \theta_0$. Then, approximately,

$$\hat{\theta} - \theta_0 = (X'P^{-1}X)^{-1}X'P^{-1}(f - p_0)$$

and since

$$X(\hat{\theta} - \theta_0) \simeq \hat{p} - p_0$$

we have the approximation

$$\hat{p} - p_0 = X(X'P^{-1}X)^{-1}X'P^{-1}(f - p_0) \qquad (7.11)$$

Geometrically, the situation is shown in Figure 7.1. This shows the one-step situation where we require only one iteration to get from θ_0 to $\hat{\theta}$. In the general case the regression of a vector y on the columns of a matrix X is equivalent to finding the orthogonal projection of y onto the range of X, that is, the orthogonal projection onto the vector space spanned by the columns of X. In the present case, orthogonal projections are relative to the metric P^{-1}. Detailed reasons for this will be given below. Figure 7.1 shows that the orthogonal projection, relative to P^{-1} of $X_0\theta_0 + f - p_0$ onto the range of X, $\mathfrak{R}(X)$, is $X_0\theta_0 + \hat{p} - p_0$, and these are the fitted values for the regression. The figure also shows that we could formulate the regression in terms of an alternative dependent variable, $f - p_0$, whose orthogonal projection,

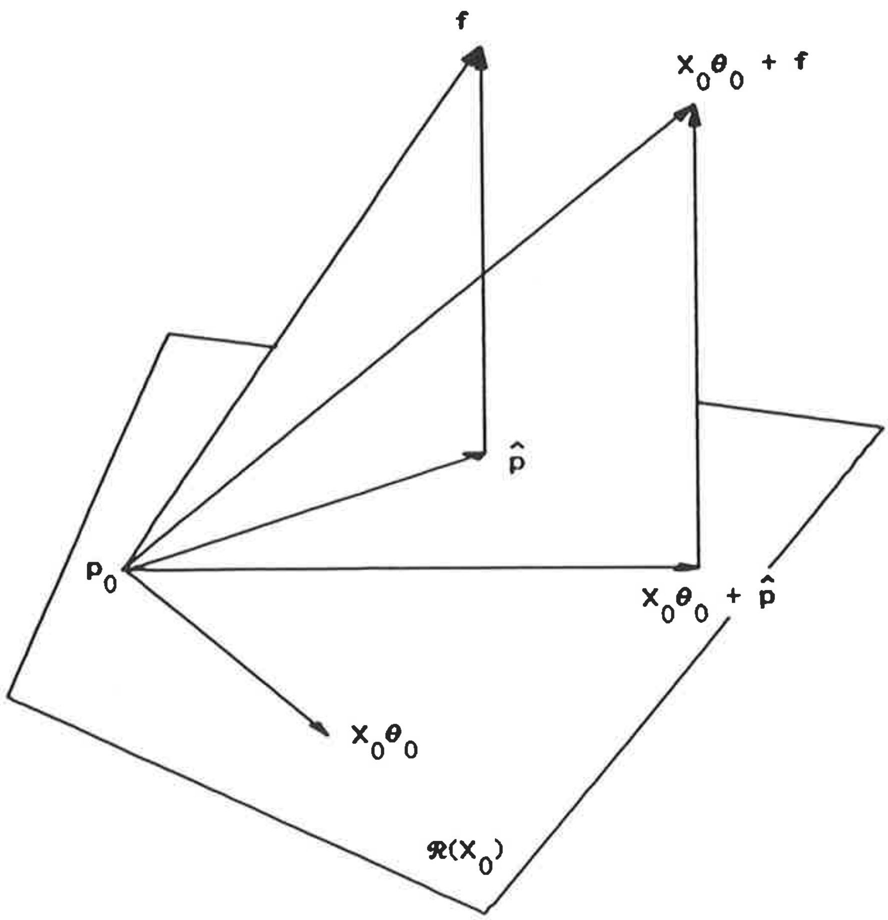

Figure 7.1 Geometry of maximum likelihood estimation as regression.

relative to $\mathbf{P}^{-1}$, onto $\mathfrak{R}(\mathbf{X})$, is $\hat{\mathbf{p}} - \mathbf{p}_0$. This formulation leads to the same estimate of $\boldsymbol{\theta}$, but use of $\mathbf{X}_0\boldsymbol{\theta}_0 + \mathbf{f} - \mathbf{p}_0$ as the dependent variable provides the estimate of $\boldsymbol{\theta}$ directly from the iterated regression.

This iterative regression formulation of the solution of the maximum likelihood equations has as its basis large sample maximum likelihood theory. Sir Ronald Fisher in his book *Statistical Methods and Scientific Inference* (3rd Edition, 1973) sought "to bring a wider class of cases into logical connection with the Analysis of Variance". The regression formulation implies that we can apply all the ideas, both mathematical and inferential, of analysis of variance, regression and covariance, to estimation.

7.5 Sample and contrast spaces

7.5.1 Sample space We shall be considering parameters of the form $\mathbf{q}^{*\prime}\mathbf{p}$ and their parametric and nonparametric estimators $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ and $\mathbf{q}^{*\prime}\mathbf{f}$. The sample relative frequencies, $\mathbf{f}$, lie in an $m-1$ dimensional hyperplane determined by $\mathbf{1}_m{}'\mathbf{f} = 1$. Since it is mathematically simpler to deal with quantities which add to zero rather than one, we subtract some fixed probability vector $\mathbf{p}_0$ from $\mathbf{f}$, and consider the equivalent problem of estimating $\mathbf{q}^{*\prime}(\mathbf{p}-\mathbf{p}_0)$ by $\mathbf{q}^{*\prime}(\hat{\mathbf{p}}-\mathbf{p}_0)$ and $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$.

The vector $\mathbf{f}-\mathbf{p}_0$, then, lies in a subspace of R^m namely $\{\mathbf{1}_m\}^{\#}$, where

$$\{\mathbf{1}_m\}^{\#} = \{\mathbf{r} \mid \mathbf{1}_m{}'\mathbf{r} = 0,\ \mathbf{r} \in R^m\}$$

We regard this subspace as the *sample space*, $\mathcal{S}$. The sample vectors $\mathbf{f}-\mathbf{p}_0$ will lie in a bounded subset of $\mathcal{S}$.

Thus the sample space includes all vectors which annihilate $\mathbf{1}_m$. This includes the vectors of derivatives of the probabilities with respect to the parameters

$$\mathbf{p}_{\theta_i} \qquad i = 1, \ldots, k.$$

Defining X as above, we thus have $\mathbf{1}_m{}'\mathbf{X} = \mathbf{0}_k{}'$.

Also in the sample space is the vector $\mathbf{p}-\mathbf{p}_0 = \mathbf{p}(\boldsymbol{\theta}) - \mathbf{p}(\boldsymbol{\theta}_0)$, since $\mathbf{1}_m{}'(\mathbf{p}-\mathbf{p}_0) = 0$. As $\boldsymbol{\theta}$ varies, $\mathbf{p}-\mathbf{p}_0$ traces out a k dimensional surface in $\mathcal{S}$, which we call the *estimation locus*.

The tangent subspace at $\mathbf{p}(\boldsymbol{\theta})$ to the estimation locus is the range of $\mathbf{X}(\boldsymbol{\theta})$, denoted by $\mathcal{R}(\mathbf{X})$ and given by

$$\mathfrak{R}(X) = \{y \mid y = Xs,\ y \in R^m,\ s \in R^k\}.$$

The vector $\hat{p} - p_0$ will lie in this subspace.

The sample space is a Euclidean or inner product space. We defer discussion of the metric of the inner product and the identity operator on sample space until we have considered the dual space in the next subsection.

The results of this section are summarised in Table 7.3.

7.5.2 <u>Contrast space</u> Let x be a vector of sample space. A linear functional on sample space is a vector $q^* \in R^{m*}$, with value the (scalar) contrast $q^{*\prime} x$. Since

$$1_m' x = 0$$

the addition to q^* of multiples of the vector 1_m does not alter the value $q^{*\prime} x$. Hence we may consider q^* as modulo 1_m in this context, and define the conjugate or dual space, which in accordance with statistical usage we call the *contrast space*, to be the quotient space of cosets $q^* + \{1_m\}$. That is, we define the contrast space as

$$\mathfrak{C}^* = R^{m*}/\{1_m\} \quad = \quad \{q^* + \{1_m\} \mid q^* \in R^{m*}\}$$

The asterisk denotes the space as the dual space, or a vector as belonging to this space.

While we can consider q^* as modulo 1_m in relation to the contrast $q^{*\prime}(p - p_0)$, when we go back to consideration of $q^{*\prime} p$ we must take note of which element of the coset $q^* + \{1_m\}$ has been used in order to obtain the correct interpretation.

Table 7.3

Summary of properties of Sample and Contrast Spaces

Sample space, $\mathcal{S}$	Contrast space, $\mathcal{C}^*$
$\{1_m\}^\# \subset R^m$	$R^{m*}/\{1_m\}$
elements	
$f - p_0$	
$p - p_0$	
$\hat{p} - p_0$	
p_{θ_i}	$P^{-1}p_{\theta_i} + \{1_m\}$
	(the score-functionals)
Estimation locus, $\subset \{1_m\}^\#$	
elements	
$p - p_0$	
tangent subspace	**score-functional subspace**
$\mathfrak{R}(X)$	$\mathfrak{R}(P^{-1}X)$ mod $\{1_m\}$
inner product matrix	
$\Sigma^- = P^{-1} - 1_m 1_m'$	$\Sigma = P - pp'$
(restricted to $\{1_m\}^\#$)	
inner product	
$\langle u,v\rangle = u'\Sigma^- v$	$\langle u^*,v^*\rangle = u^{*\prime}\Sigma v^*$
identity operator	
$I_m - p1_m'$	$I_m - 1_m p'$
orthogonal decomposition	
$\mathcal{T} = \mathfrak{R}(X)$, dimension k	$\mathcal{U}^\# = \mathfrak{R}(P^{-1}X)$, dimension k
$\mathcal{U} = \mathfrak{R}(\Sigma L)$, dim. $m-k-1$	$\mathcal{T}^\# = \mathfrak{R}(L)$ mod $\{1_m\}$, dim $m-k-1$
$\mathcal{T} \perp \mathcal{U}$ relative to Σ^-	$\mathcal{T}^\# \perp \mathcal{U}^\#$ relative to Σ

7.5.3 <u>Inner product metrics and identity operators</u> Both the sample and the contrast space have natural inner products, and thus an inner product

$$\langle u , v \rangle = u'Mv$$

is defined for any two vectors **u, v** in the space. **M** is the matrix of the inner product, or the metric of the space.

The matrix

$$\boldsymbol{\Sigma} = \mathbf{P} - \mathbf{p}\mathbf{p}' \tag{7.12}$$

forms the natural metric on contrast space, since the covariance of two contrasts $\mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ and $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ is given by the bilinear form

$$\frac{1}{n}\,\mathbf{q}_1^{*\prime}\,\boldsymbol{\Sigma}\,\mathbf{q}_2^{*}.$$

Therefore, as the metric on sample space, we can take a generalised inverse $\boldsymbol{\Sigma}^-$ of $\boldsymbol{\Sigma}$ (Dempster, 1969). We choose

$$\boldsymbol{\Sigma}^- = \mathbf{P}^{-1} - \mathbf{1}_m\mathbf{1}_m'. \tag{7.13}$$

To see the reasons for this choice, consider the representation of the multinomial distribution as the joint distribution of m independent Poisson variates, conditional on their sum (e.g. Fisher, 1922; Rao, 1952). The unconditional variance matrix of these variates is $n\mathbf{P}$, and we have (by 7.12) the decomposition

$$n\mathbf{P} = n\boldsymbol{\Sigma} + n\,\mathbf{p}\mathbf{p}' \tag{7.14}$$

Corresponding to this we can decompose the information matrix of the Poisson variates, $\frac{1}{n}\mathbf{P}^{-1}$, into two parts which are respectively generalised inverses of the components of (7.14) and which annihilate the other component:

$$\frac{1}{n}\mathbf{P}^{-1} = \frac{1}{n}\,\boldsymbol{\Sigma}^- + \frac{1}{n}\,\mathbf{1}_m\mathbf{1}_m'.$$

This leads to (7.13) as the choice of metric on the sample space. Instead of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^-$, we could equally use $n\boldsymbol{\Sigma}$ and $\frac{1}{n}\boldsymbol{\Sigma}^-$, but it is convenient to use the unscaled versions.

The choice of these two metrics (7.12) and (7.13) leads naturally to the choice of matrices which represent most conveniently the identity operators on the two spaces. On the sample space we choose

$$\Sigma\Sigma^{-} = I_m - p1_m'$$

since $\mathfrak{R}(\Sigma\Sigma^{-})$ is the sample space, and $\Sigma\Sigma^{-}$ leaves vectors in the sample space invariant and maps m-component vectors outside the sample space into that space. On the contrast space, we choose

$$\Sigma^{-}\Sigma = I_m - 1_m p'$$

which is the transpose of the sample space identity. $\Sigma^{-}\Sigma$ maps m-component vectors into a cross-section of contrast space.

7.5.4 <u>The score-functional subspace of contrast space</u> In general the contrast space contains mappings by Σ^{-} of the vectors of the sample space, including the score-functionals (that is, the functionals giving the components of the score vector)

$$\Sigma^{-}p_{\theta_i} = P^{-1}p_{\theta_i}$$

which are the columns of $P^{-1}X$. The score-functionals span the *score-functional subspace* $\mathfrak{R}(P^{-1}X)$ mod $\{1_m\}$.

7.5.5 <u>Orthogonal decompositions of sample and contrast spaces</u> Let $\mathcal{T}$ be a k dimensional subspace of the sample space, and $\mathcal{T}^{\#}$ its $m-k-1$ dimensional annihilator in contrast space. The columns of a matrix X ($m \times k$) will be a *basis* of $\mathcal{T}$ if they lie in $\mathcal{T}$, X has rank k and $1_m'X = 0$. Then $\mathcal{T} = \mathfrak{R}(X)$.

Similarly the cosets $l_i^* + \{1_m\}$, $i = 1, \ldots, m-k-1$ will be a basis of $\mathcal{T}^{\#}$ if the l_i^* are the columns of a matrix L ($m \times (m-k-1)$) such that $L'X = 0$ and $[L|1_m]$ has rank $(m-k)$. Then $\mathcal{T}^{\#} = \mathfrak{R}(L)$ mod$\{1_m\}$.

The variance matrix Σ maps $\mathcal{T}^{\#}$ into the orthogonal complement $\mathcal{U}$ of $\mathcal{T}$ in sample space, relative to the metric $\Sigma^{-} = P^{-1} - 1_m 1_m'$. $\mathcal{U}$ will also have dimension $m-k-1$. For conciseness and emphasis, we say $\mathcal{U}$ is Σ^{-}-orthogonal to $\mathcal{T}$. We prove this in the following lemma:

<u>Lemma</u>: *$\mathcal{T}$ and $\mathcal{U}$ are Σ^{-}-orthogonal subspaces of the sample space.*

<u>Proof</u>: By definition, $\mathcal{T} = \mathfrak{R}(X)$, $\mathcal{T}^{\#} = \mathfrak{R}(L) \bmod \{1_m\}$ and $\mathcal{U} = \mathfrak{R}(\Sigma L)$. Then for any $x \in \mathcal{T}$ and $y \in \mathcal{U}$, we have

$$x = Xw \qquad w \in R^{k}$$

$$y = \Sigma Lz \qquad z \in R^{m-k-1}$$

Therefore

$$\begin{aligned} \langle x , y \rangle = x'\Sigma^{-}y &= w'X'P^{-1}\Sigma Lz - w'X'1_m 1_m'\Sigma Lz \\ &= w'X'Lz - w'X'1_m p'Lz - 0 \\ &= 0 \end{aligned}$$

which proves the lemma. □

Thus we have the decomposition of the sample space into the direct sum of Σ^{-}-orthogonal subspaces

$$\begin{aligned} \{1_m\}^{\#} &= \mathcal{T} \oplus \mathcal{U} \\ &= \mathfrak{R}(X) \oplus \mathfrak{R}(X)^{\perp} \\ &= \mathfrak{R}(X) \oplus \mathfrak{R}(\Sigma L) \end{aligned}$$

An equivalent situation holds in the contrast space. Σ^{-} maps $\mathcal{T}$ into $\mathcal{U}^{\#}$, the k dimensional annihilator of $\mathcal{U}$. Then

$$\mathcal{U}^{\#} = \mathfrak{R}(\Sigma^{-} X) = \mathfrak{R}(P^{-1} X)$$

$\mathcal{U}^{\#}$ and $\mathcal{T}^{\#}$ are Σ-orthogonal complements of contrast space. This can be shown by a proof analogous to that of the lemma above.

Then in the contrast space, we have the decomposition into Σ-orthogonal subspaces

$$\begin{aligned} R^{m*}/\{\mathbf{1}_m\} &= \mathcal{U}^{\#} \oplus \mathcal{T}^{\#} \\ &= \mathfrak{R}(P^{-1}X) \oplus \mathfrak{R}(P^{-1}X)^{\perp} \\ &= \mathfrak{R}(P^{-1}X) \oplus \mathfrak{R}(L) \bmod \{\mathbf{1}_m\} \end{aligned}$$

These decompositions are summarised in Table 7.3

Since the cosets $\mathbf{l}_i^* + \{\mathbf{1}_m\}$ are a basis of $\mathcal{T}^{\#}$ the columns of $\mathbf{\Sigma L}$ will be a basis of $\mathcal{U}$. Additionally we have the columns of $\mathbf{X}$ as a basis of $\mathcal{T}$. Then since $\mathcal{T}$ and $\mathcal{U}$ are orthogonal complements in sample space, the columns of $\mathbf{X}$ and $\mathbf{\Sigma L}$ will be a basis of sample space. Similarly the columns of $\mathbf{P}^{-1}\mathbf{X}$ and $\mathbf{L}$ will be a basis of contrast space.

We note also that the inclusion of $\mathbf{p}$ with the basis for the sample space gives a basis for R^m; similarly, the inclusion of $\mathbf{1}_m$ with the basis of contrast space gives a basis of R^{m*}.

7.6 Decomposition theorem

In some sections of this chapter, we wish to decompose vectors in sample space into components in $\mathcal{T}$ and $\mathcal{U}$, or those in contrast space into components in $\mathcal{T}^{\#}$ and $\mathcal{U}^{\#}$. We therefore derive the decompositions of the identity transforms of the sample and contrast spaces into projections onto the relevant subspaces. Similarly we will need decompositions of the metrics of the two spaces, the variance and information matrices. These decompositions are given in the following theorem.

<u>Theorem</u>: *Let $\mathcal{T}$ be a* k*-dimensional subspace of sample space, and $\mathcal{U}$ its* m−k−1 *dimensional Σ^{-}-orthogonal complement. Let $\mathcal{T}^{\#}$ be the* m−k−1 *dimensional subspace of contrast space which is the annihilator of $\mathcal{T}$, and $\mathcal{U}^{\#}$ the* k *dimensional Σ-orthogonal complement of $\mathcal{U}$ in contrast space. Relative to $\mathcal{T}$ and $\mathcal{U}$ in sample space, and $\mathcal{T}^{\#}$ and $\mathcal{U}^{\#}$ in contrast space, we have the following decompositions*

1. *of the identity transform in sample space into idempotents*

$$I_m - p1_m' = X(X'P^{-1}X)^{-1}X'P^{-1} + \Sigma L(L'\Sigma L)^{-1}L'(I_m - p1_m') \qquad (7.15)$$

2. *of the identity transform in contrast space into idempotents*

$$I_m - 1_m p' = P^{-1}X(X'P^{-1}X)^{-1}X' + (I_m - 1_m p')\, L(L'\Sigma L)^{-1}L'\Sigma \qquad (7.16)$$

3. *of the variance matrix*

$$\Sigma = X(X'P^{-1}X)^{-1}X' + \Sigma L(L'\Sigma L)^{-1}L'\Sigma \qquad (7.17)$$

4. *of the information matrix*

$$P^{-1} - 1_m 1_m' = P^{-1}X(X'P^{-1}X)^{-1}X'P^{-1} + (I_m - 1_m p')\, L(L'\Sigma L)^{-1}L'\,(I_m - p1_m') \qquad (7.18)$$

<u>Proof</u>: All four decompositions can be proved individually. However having derived any one decomposition, the others follow from it simply by appropriate pre- and/or post-multiplication. We shall do this and subsequently indicate briefly how direct proofs can be given.

We consider the decomposition of the identity transform on sample space

$$I_m - p1_m'$$

into projections onto the subspaces $\mathcal{T} = \mathfrak{R}(X)$ and $\mathcal{U} = \mathfrak{R}(\Sigma L)$. We suppose

$$I_m - p1_m' = E_1 + E_2 \tag{7.19}$$

Then E_1 is the projection onto $\mathcal{T}$ parallel to $\mathcal{U}$, that is to say, the projection whose range is $\mathfrak{R}(X)$ and whose kernel is $\mathfrak{R}(X)^{\perp}$, where the perpendicularity is relative to the metric $\Sigma^- = P^{-1} - 1_m 1_m'$. One form which satisfies these requirements is

$$\begin{aligned} E_1 &= XAX'\Sigma^- \\ &= XAX'P^{-1} \end{aligned}$$

where A is a $k \times k$ matrix to be determined. Since E_1 is a projection matrix, it is idempotent:

$$XAX'P^{-1}XAX'P^{-1} = XAX'P^{-1}$$

that is

$$AX'P^{-1}XA = A$$

Since E_1 must have rank k (= dimension of $\mathfrak{R}(X)$), A must be nonsingular. Pre- and post-multiplying by A^{-1}, and inverting, gives

$$A = (X'P^{-1}X)^{-1}$$

Therefore

$$E_1 = X(X'P^{-1}X)^{-1}X'P^{-1} \tag{7.20}$$

is the first required component of the identity transform.

Similarly we can deduce the form of E_2. E_2 must be the projection whose range is $\mathfrak{R}(\Sigma L)$ and whose kernel is $\mathfrak{R}(\Sigma L)^{\perp}$, again with perpendicularity defined relative to the metric Σ^{-}. Therefore we can take

$$E_2 = \Sigma LAL'\Sigma(P^{-1} - 1_m 1_m')$$
$$= \Sigma LAL'(I_m - p1_m')$$

Since E_2 is a projection, it is idempotent:

$$\Sigma LAL'(I_m - p1_m')\Sigma LAL'(I_m - p1_m') = \Sigma LAL'(I_m - p1_m')$$

that is

$$AL'(I_m - p1_m')\Sigma LA = A$$

which reduces to

$$AL'\Sigma LA = A .$$

Now ΣL has full column rank because L is carefully constructed to have no linear function of its columns entirely in the null space of Σ. Hence $L'\Sigma L$ is nonsingular, and we can write

$$A = (L'\Sigma L)^{-1}$$

Therefore

$$E_2 = \Sigma L(L'\Sigma L)^{-1}L'(I_m - p1_m') \tag{7.21}$$

is the second component of the identity transform. Thus equations (7.19), (7.20) and (7.21) give the first decomposition (7.15) of the theorem:

$$I_m - p1_m' = X(X'P^{-1}X)^{-1}X'P^{-1} + \Sigma L(L'\Sigma L)^{-1}L'(I_m - p1_m')$$

The other decompositions then follow from this one as follows:

i. *the identity transform in contrast space* (7.16): pre-multiply by P^{-1} and post-multiply by P.

ii. *the variance matrix* (7.17): post-multiply by $\mathbf{P}$.

iii. *the information matrix* (7.18): pre-multiply by $\mathbf{P}^{-1}$.

This completes the proof of the theorem. □

An indication of the method of direct proof of the other three decompositions follows.

i. To deduce a decomposition of the identity in contrast space, we proceed similarly to the proof above. For a supposed decomposition

$$\mathbf{I}_m - \mathbf{1}_m\mathbf{p}' = \mathbf{E}_1^* + \mathbf{E}_2^*$$

we note that $\mathbf{E}_1^*$ must be the projection whose range is $\mathcal{R}(\mathbf{P}^{-1}\mathbf{X})$ and whose kernel is $\mathcal{R}(\mathbf{P}^{-1}\mathbf{X})^{\perp}$, while $\mathbf{E}_2^*$ is the projection whose range is $\mathcal{R}(\mathbf{L})$ and whose kernel is $\mathcal{R}(\mathbf{L})^{\perp}$. We then follow a similar argument to that for the sample space.

ii. To deduce the decomposition of the variance matrix, $\boldsymbol{\Sigma}$, we can use equation (7.19) to decompose a vector $\mathbf{x}$ of sample space into components $\mathbf{E}_1\mathbf{x}$ and $\mathbf{E}_2\mathbf{x}$. Since $\mathbf{E}_1$ and $\mathbf{E}_2$ are projections on $\boldsymbol{\Sigma}^-$-orthogonal complements, consideration of the variances of $\mathbf{x}$ and its two components gives, say,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$$

We then have

$$\boldsymbol{\Sigma}_1 = V(\mathbf{E}_1\mathbf{x}) = \mathbf{E}_1\boldsymbol{\Sigma}\mathbf{E}_1'$$

$$\boldsymbol{\Sigma}_2 = V(\mathbf{E}_2\mathbf{x}) = \mathbf{E}_2\boldsymbol{\Sigma}\mathbf{E}_2'$$

and substitution for $\mathbf{E}_1$ and $\mathbf{E}_2$ from (7.20) and (7.21) gives the required decomposition.

For this decomposition, we note that

a. rank (Σ_1) = k, rank (Σ_2) = $m-k-1$

b. $\Re(\Sigma_1) = \Re(X)$, $\Re(\Sigma_2) = \Re(\Sigma L)$,

c. Σ^- is a generalised inverse of Σ_1, since

$$\begin{aligned}\Sigma_1\Sigma^-\Sigma_1 &= X(X'P^{-1}X)^{-1}X'(P^{-1} - 1_m1_m')X(X'P^{-1}X)^{-1}X' \\ &= X(X'P^{-1}X)^{-1}X' \\ &= \Sigma_1\end{aligned}$$

d. Similarly, Σ^- is a generalised inverse of Σ_2.

iii. To decompose the information matrix, we note that the information matrix is a generalised inverse of the variance matrix, and use the decomposition of the variance matrix already derived:

$$\begin{aligned}\Sigma^- &= \Sigma^-\Sigma\Sigma^- \\ &= \Sigma^-(\Sigma_1 + \Sigma_2)\Sigma^- \\ &= \Sigma_1^- + \Sigma_2^-\end{aligned}$$

Substitution for Σ_1 and Σ_2 produces the required decomposition.

Equivalent decompositions for the multinormal distribution, with Σ of full rank, were given by James (1973).

7.7 Partitions of contrasts in parametric estimation

7.7.1 <u>Introduction</u> A contrast, that is, a linear function of the class probabilities, can be used to highlight aspects of inference for grouped frequency distributions. In a previous section we have discussed some linear functionals which generate contrasts of interest in the present study of alcohol consumption.

Typically in the estimation of contrasts from a set of data, one specification is fitted and then all contrasts estimated using this specification. We suggest that the assumptions made in using any specification are never perfectly satisfied, but may be better satisfied for some contrasts than for others, and that the validity of assumptions made in estimating a contrast needs to be examined for each contrast estimated.

As we have explained earlier, for mathematical convenience we subtract some fixed probability vector, $\mathbf{p}_0$, from $\mathbf{p}$, and consider contrasts of the form $\mathbf{q}^{*\prime}(\mathbf{p}-\mathbf{p}_0)$ rather than $\mathbf{q}^{*\prime}\mathbf{p}$. This leaves the variance of $\mathbf{q}^{*\prime}\mathbf{p}$ unchanged.

In this section, we will be using the decomposition theorem of the previous section to partition the nonparametric estimate

$$\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$$

of the contrast

$$\mathbf{q}^{*\prime}(\mathbf{p}-\mathbf{p}_0)$$

into two components, one representing the parametric estimate

$$\mathbf{q}^{*\prime}(\hat{\mathbf{p}}-\mathbf{p}_0)$$

of the contrast, the second component being an estimator of zero if we can assume that the parametric specification is correct.

By similarly decomposing the variance and information matrices, we can get an indication of the extent to which the estimate of the contrast is dependent on the values of the parameters, and the extent to which it is dependent on the choice of model.

We begin by examining the partition of contrasts.

7.7.2 <u>Partitions of contrasts</u> We can approach this in two ways: via decomposition of the sample space vector $(\mathbf{f}-\mathbf{p}_0)$, or via the decomposition of the linear functional $\mathbf{q}^*$ of contrast space. We will show that both approaches lead to identical results.

a. <u>Sample space approach.</u> The identity

$$\mathbf{f}-\mathbf{p}_0 = (\hat{\mathbf{p}}-\mathbf{p}_0) + (\mathbf{f}-\hat{\mathbf{p}})$$

by the nature of the estimation process, decomposes the sample space vector $\mathbf{f}-\mathbf{p}_0$ into its projection on the tangent subspace $\mathfrak{R}(X)$ and its Σ^--orthogonal complement, $\mathfrak{R}(X)^\perp$ or $\mathfrak{R}(\Sigma L)$.

We can also effect an orthogonal decomposition of $\mathbf{f}-\mathbf{p}_0$ using the decomposition of the sample space identity transform derived in the previous section (equation (7.15)):

$$(I_m - \mathbf{p}\mathbf{1}_m')(\mathbf{f}-\mathbf{p}_0) = E_1(\mathbf{f}-\mathbf{p}_0) + E_2(\mathbf{f}-\mathbf{p}_0)$$

that is,

$$\mathbf{f}-\mathbf{p}_0 = E_1(\mathbf{f}-\mathbf{p}_0) + E_2(\mathbf{f}-\mathbf{p}_0) \qquad (7.22)$$

where, by (7.20) and (7.21)

$$E_1 = X(X'P^{-1}X)^{-1}X'P^{-1}$$

$$E_2 = \Sigma L(L'\Sigma L)^{-1}L'(I_m - p1_m') \ .$$

Thus

$$E_1(f - p_0) = X(X'P^{-1}X)^{-1}X'P^{-1}(f - p_0)$$

which are the one-step approximations to the fitted values (7.11) from the regression formulation of maximum likelihood estimation. Thus

$$E_1(f - p_0) = \hat{p} - p_0 \ .$$

Also we can write

$$\begin{aligned} E_2(f - p_0) &= \{(I_m - p1_m') - E_1\} (f - p_0) \\ &= f - \hat{p} \end{aligned}$$

Thus the two decompositions of $f - p_0$ are equivalent, and, in summary, we have

$$f - p_0 = (\hat{p} - p_0) + (f - \hat{p})$$

$$f - p_0 = E_1(f - p_0) + E_2(f - p_0)$$

The first component lies in the tangent subspace $\mathfrak{R}(X)$, while the second lies in its Σ^{-}-orthogonal complement.

We can then use these decompositions to partition the contrast $q^{*\prime}(f - p_0)$:

$$q^{*\prime}(f - p_0) = q^{*\prime}(p - p_0) + q^{*\prime}(f - \hat{p}) \qquad (7.23)$$

$$q^{*\prime}(f - p_0) = q^{*\prime}E_1(f - p_0) + q^{*\prime}E_2(f - p_0) \qquad (7.24)$$

We now show that we can arrive at the same result by decomposing the linear functional q^*.

b. <u>Contrast space approach.</u> Since q^* lies in the contrast space, we can use the decomposition of the contrast space transform (7.16) to decompose q^*:

$$(\mathbf{I}_m - \mathbf{1}_m\mathbf{p}')\mathbf{q}^* = \mathbf{E}_1^*\mathbf{q}^* + \mathbf{E}_2^*\mathbf{q}^*$$

where

$$\begin{aligned} \mathbf{E}_1^* &= \mathbf{P}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{E}_1' \\ \mathbf{E}_2^* &= (\mathbf{I}_m - \mathbf{1}_m\mathbf{p}')\mathbf{L}(\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L})^{-1}\mathbf{L}'\boldsymbol{\Sigma} \\ &= \mathbf{E}_2' \end{aligned}$$

The first component is in the score-functional subspace, $\mathfrak{R}(\mathbf{P}^{-1}\mathbf{X})$, while the second component lies in its $\boldsymbol{\Sigma}$-orthogonal complement, $\mathfrak{R}(\mathbf{L})$ mod $\{\mathbf{1}_m\}$.

We can then apply this decomposition of $\mathbf{q}^*$ to $(\mathbf{f} - \mathbf{p}_0)$:

$$\{(\mathbf{I}_m - \mathbf{1}_m\mathbf{p}')\mathbf{q}^*\}'(\mathbf{f} - \mathbf{p}_0) = (\mathbf{E}_1^*\mathbf{q}^*)'(\mathbf{f} - \mathbf{p}_0) + (\mathbf{E}_2^*\mathbf{q}^*)'(\mathbf{f} - \mathbf{p}_0)$$

which yields

$$\begin{aligned} \mathbf{q}^{*\prime}(\mathbf{f} - \mathbf{p}_0) &= \mathbf{q}^{*\prime}\mathbf{E}_1^{*\prime}(\mathbf{f} - \mathbf{p}_0) + \mathbf{q}^{*\prime}\mathbf{E}_2^{*\prime}(\mathbf{f} - \mathbf{p}_0) \\ &= \mathbf{q}^{*\prime}\mathbf{E}_1(\mathbf{f} - \mathbf{p}_0) + \mathbf{q}^{*\prime}\mathbf{E}_2(\mathbf{f} - \mathbf{p}_0) \end{aligned}$$

which is equation (7.24). Thus the approaches from the sample and contrast spaces are equivalent.

We put

$$\begin{aligned} \mathbf{q}_1^* &= \mathbf{E}_1^*\mathbf{q}^* = \mathbf{E}_1'\mathbf{q}^* \\ \mathbf{q}_2^* &= \mathbf{E}_2^*\mathbf{q}^* = \mathbf{E}_2'\mathbf{q}^* \end{aligned}$$

and have, finally,

$$\mathbf{q}^{*\prime}(\mathbf{f} - \mathbf{p}_0) = \mathbf{q}^{*\prime}(\hat{\mathbf{p}} - \mathbf{p}_0) + \mathbf{q}^{*\prime}(\mathbf{f} - \hat{\mathbf{p}}) \quad (7.25)$$

$$= \mathbf{q}_1^{*\prime}(\mathbf{f} - \mathbf{p}_0) + \mathbf{q}_2^{*\prime}(\mathbf{f} - \mathbf{p}_0) \quad (7.26)$$

Now $\mathbf{q}^{*\prime}(\mathbf{f} - \mathbf{p}_0)$ is the nonparametric estimate of the contrast $\mathbf{q}^{*\prime}(\mathbf{p} - \mathbf{p}_0)$, and we have decomposed it into the parametric estimate

$$\mathbf{q}^{*\prime}(\hat{\mathbf{p}} - \mathbf{p}_0) = \mathbf{q}_1^{*\prime}(\mathbf{f} - \mathbf{p}_0)$$

and a second component

$$\mathbf{q}^{*\prime}(\mathbf{f}-\hat{\mathbf{p}}) = \mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$$

whose expectation is zero if we can assume that the parametric specification is correct.

7.7.3 Partitions of χ^2 In the same manner as we partitioned the contrast $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$, we can use the decomposition theorem to partition χ^2. For χ^2 is the quadratic form in $\mathbf{y} = \mathbf{f}-\mathbf{p}_0$ of the information matrix

$$n\Sigma^- = n(\mathbf{P}^{-1} - \mathbf{1}_m\mathbf{1}_m')$$

that is

$$n(\mathbf{f}-\mathbf{p}_0)'\mathbf{P}^{-1}(\mathbf{f}-\mathbf{p}_0) \sim \chi^2_{m-1}$$

Using equation (7.18) we can partition this into two components

$$n(\mathbf{f}-\mathbf{p}_0)'\mathbf{P}^{-1}(\mathbf{f}-\mathbf{p}_0) = n(\mathbf{f}-\mathbf{p}_0)'\,\mathbf{I}_1\,(\mathbf{f}-\mathbf{p}_0) + n(\mathbf{f}-\mathbf{p}_0)'\,\mathbf{I}_2\,(\mathbf{f}-\mathbf{p}_0) \quad (7.27)$$

where

$$\mathbf{I}_1 = \mathbf{P}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}^{-1}$$
$$\mathbf{I}_2 = \mathbf{L}(\mathbf{L}'\Sigma\mathbf{L})^{-1}\mathbf{L}'$$

with evaluation being at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. We then have a partition of χ^2_{m-1} into two components. The first part, on k degrees of freedom, is a test of the deviations of the parameter values from the maximum likelihood estimates, and, at $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$, will be zero (since $\mathbf{X}'\mathbf{P}^{-1}\mathbf{f} = \mathbf{0}_k$ are the maximum likelihood equations (7.05)). The second component, on $m-k-1$ degrees of freedom, is the usual χ^2 goodness-of-fit test statistic, testing deviations of the data from the model. This partition of χ^2 was given by Fisher (1963).

However this goodness-of-fit test is for the overall fit of the model, and is not specific to any contrast. By considering partitions of the variance matrix, we can derive tests of goodness-of-fit which relate more to particular contrasts.

7.7.4 Partitions of variance The variances of the components of the decomposition (7.22) of $\mathbf{f}-\mathbf{p}_0$ are given by the decomposition of the variance matrix $\boldsymbol{\Sigma}$ from equation (7.17), that is

$$V(\mathbf{f}-\mathbf{p}_0) = V(E_1(\mathbf{f}-\mathbf{p}_0)) + V(E_2(\mathbf{f}-\mathbf{p}_0))$$

or

$$\frac{1}{n}\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{\Sigma}_1 + \frac{1}{n}\boldsymbol{\Sigma}_2$$

where

$$\boldsymbol{\Sigma}_1 = \mathbf{X}(\mathbf{X}'\mathbf{P}^{-1}\mathbf{X})^{-1}\mathbf{X}'$$

$$\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}\mathbf{L}(\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L})^{-1}\mathbf{L}'\boldsymbol{\Sigma}$$

Thus the variances of the contrast $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ and its components are

$$V(\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)) = V(\mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0)) + V(\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0))$$

$$\frac{1}{n}\mathbf{q}^{*\prime}\,\boldsymbol{\Sigma}\,\mathbf{q}^* = \frac{1}{n}\mathbf{q}^{*\prime}\,\boldsymbol{\Sigma}_1\,\mathbf{q}^* + \frac{1}{n}\mathbf{q}^{*\prime}\,\boldsymbol{\Sigma}_2\,\mathbf{q}^* \qquad (7.28)$$

which some simple algebra shows is equal to

$$= \frac{1}{n}\mathbf{q}_1^{*\prime}\boldsymbol{\Sigma}\mathbf{q}_1^* + \frac{1}{n}\mathbf{q}_2^{*\prime}\boldsymbol{\Sigma}\mathbf{q}_2^* \qquad (7.29)$$

The covariance of $\mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ and $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ is $\frac{1}{n}\mathbf{q}_1^{*\prime}\boldsymbol{\Sigma}\,\mathbf{q}_2^*$, which is zero since $\mathbf{q}_1^*$ and $\mathbf{q}_2^*$ lie in $\boldsymbol{\Sigma}$-orthogonal subspaces of contrast space.

The two forms (7.28) and (7.29) correspond to the two approaches taken above to the partition of $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$, namely via the sample and contrast spaces, and are likewise equivalent.

7.7.5 Example Let us illustrate the foregoing ideas using the Busselton data. Suppose our interest lies in the true mean consumption, and in the average consumption per head in excess of 60 g/day. Then if we are unwilling to make any assumption about the form of the distribution of consumption, our estimates are the nonparametric contrasts

$$\text{mean consumption} \qquad \mathbf{q}_t^{*\prime}\mathbf{f} = 15.415 \text{ g/day}$$

and

$$\text{excess consumption} \qquad \mathbf{q}_e^{*\prime}\mathbf{f} = 0.250 \text{ g/day}$$

where $\mathbf{q}_t^*$ and $\mathbf{q}_e^*$ are as defined in Table 7.2.

We may however assume a parametric specification of the distribution, perhaps because our belief is that the distribution is essentially "smooth". A two parameter lognormal distribution fitted to the data results in

$$\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}) = (2.2717, 0.8529), \quad \chi_3^2 = 4.64, \quad (P = 0.20)$$

Then our estimates of the mean and excess consumption are respectively

$$\mathbf{q}_t^{*\prime}\hat{\mathbf{p}} = 15.447 \text{ g/day}$$

and

$$\mathbf{q}_e^{*\prime}\hat{\mathbf{p}} = 0.359 \text{ g/day}.$$

The question arises as to what effect our assumption of the two parameter lognormal distribution has had on the estimates of the true values. Using the foregoing theory, we can examine this by looking at the variances of the components of the contrasts.

The mathematical framework we have set up has been, for mathematical convenience, couched in terms of $\mathbf{f}-\mathbf{p}_0$ and $\hat{\mathbf{p}}-\mathbf{p}_0$ rather than $\mathbf{f}$ and $\hat{\mathbf{p}}$. However since the subtraction of the constant vector $\mathbf{p}_0$ from $\mathbf{f}$ and $\hat{\mathbf{p}}$ leaves the variances of $\mathbf{q}^{*\prime}\mathbf{f}$ and $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ unchanged, we can partition the contrasts and their variances in accordance with equations (7.26) and (7.28) respectively. This will give us two components, one concerned with the difference of the parameter values from the maximum likelihood estimates, and the other concerned with the difference between the data and the model.

Since our interest lies in parameter values at the maximum likelihood estimate, we take the fixed parameter θ_0 to be at the maximum likelihood estimate, and evaluate the partitions at $p_0 = \hat{p}$. Table 7.4 gives the results.

Table 7.4

Components of $q^{*\prime}(f-p_0)$, evaluated at $p_0 = \hat{p}$, for two contrasts

Busselton data, two parameter lognormal model
(values in parentheses are variances x 1001)

q^*	$q^{*\prime}(f-p_0)$	$q_1^{*\prime}(f-p_0)$	$q_2^{*\prime}(f-p_0)$
(10 30 50 70 90 110)′ "mean consumption"	−0.032 (169.948)	0.0 (169.690)	−0.032 (0.258)
(0 0 0 10 30 50)′ "excess consumption"	−0.109 (11.816)	0.0 (8.675)	−0.109 (3.141)

Naturally, the first component

$$q_1^{*\prime}(f-p_0)$$

of both contrasts is zero, since our evaluation is at $p_0 = \hat{p}$, while the second component

$$q_2^{*\prime}(f-p_0)$$

has zero expectation, assuming an adequate model. This represents the discrepancy between the nonparametric and parametric estimates of the contrast; that is, the effect of the assumptions of the model. For mean consumption, this estimator of zero has a value of −0.032, while for excess consumption, its value is more than three times this.

Let us now look at the variances of the components. For the estimate of mean consumption, only 0.15% of the variance of the nonparametric estimate is ascribable to deviations of the model from the data, but for the estimate of excess consumption, more than 26% of the variance of $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ is associated with this component.

While the assumption of the two parameter lognormal distribution may be adequate for the estimation of the mean consumption, we must conclude that its use for the estimation of excess consumption is less reliable.

7.7.6 Discussion In this section, we have shown that the nonparametric estimate of a contrast can be partitioned into the parametric estimate and a second component whose expectation is zero if we can assume that the parametric specification is correct. We have achieved this by decomposing the linear functional $\mathbf{q}^*$ into components lying respectively in the Σ-orthogonal subspaces $\mathcal{R}(P^{-1}X)$ and $\mathcal{R}(P^{-1}X)^{\perp}$ of contrast space.

Suppose that the specification can be assumed to be only partly correct. For example, when we make inferences on the upper tail of the distribution, as we did with the excess consumption contrast in the example above, we may be prepared to assume that the distribution is "smooth" and has an upper tail which can be reasonably graduated by the lognormal distribution, but we may not be prepared to rely on the assumption that the distribution is strictly symmetrical on the logarithmic scale.

Then a component of $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ may have non-zero expectation, and hence a component of $\mathbf{q}_2^*$ should be transferred to the part $\mathbf{q}_1^*$ of $\mathbf{q}^*$ corresponding to the parametric specification. That is, a component of $\mathbf{q}_2^*$ should be transferred from $\mathcal{R}(P^{-1}X)^{\perp}$ to $\mathcal{R}(P^{-1}X)$. We would then be more confident that for the new reduced $\mathbf{q}_2^*$, the expectation of $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ would be zero.

In the next section, we will show that this can be done either by introducing the third parameter to the two parameter lognormal distribution, or, almost equivalently, censoring the lower tail.

7.8 Modifications of the two parameter lognormal distribution: a comparison of adding a third parameter and censoring the lower tail

7.8.1 Introduction In parametric inference based on the two parametric lognormal distribution the inferences about the upper tail are influenced by the frequencies in the lower tail, but one may have good reason to think that these are essentially irrelevant. One may look for a mode of inference which releases inferences about the upper tail from evidence in the lower tail which is irrelevant to them.

We have previously noted various possible modifications to the two parameter lognormal distribution which might overcome this problem of spurious information in the lower tail influencing the fit in the upper tail. In this section, we consider two of them, namely, adding a third, threshold, parameter to the specification, and censoring the lower tail. We demonstrate quantitatively that the spurious information is removed by these methods.

We have shown that maximum likelihood estimation is equivalent to iterated weighted regression. Thus fitting the two parameter model is equivalent to regressing $\mathbf{X}\boldsymbol{\theta} + \mathbf{f} - \mathbf{p}$ on $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$ with $\mathbf{P}^{-1}$ as weight matrix. This uses two degrees of freedom out of the total of $m-1$ available.

Fitting the three parameter distribution adds a third independent variable into the regression, namely $\mathbf{p}_\tau$ orthogonalised to $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$. This is equivalent to fitting $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$ with covariance on $\mathbf{p}_\tau$. The covariate uses an additional one degree of freedom.

We can regard fitting the censored distribution as again equivalent to regressing $\mathbf{X}\boldsymbol{\theta} + \mathbf{f} - \mathbf{p}$ on $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$ with covariance, but this time, as the covariate, we define, in a heuristic way, a vector of contrast space for each degree of freedom lost through censoring, and map it into the sample space.

Using this covariance approach we can approximate the three parameter and censored two parameter distributions in such a way that all three fitted models have the same fitted probabilities and variance matrices. We can then compare quantitatively the covariates, and the variance, associated with each fit.

This we do by partitioning the nonparametric estimator

$$\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$$

and the corresponding variance into relevant components.

7.8.2 <u>Relationship of the three parameter and censored two parameter log-normal distributions to the two parameter lognormal distribution</u> Suppose we fit a two parameter lognormal distribution, parameters μ and σ. As we showed in Section 7.4.3, this is equivalent to iterating the regression of $\mathbf{X}\boldsymbol{\theta}+\mathbf{f}-\mathbf{p}$ on $\mathbf{X}$, with $\mathbf{P}^{-1}$ as weight matrix. In this case, the matrix $\mathbf{X}$ is the $m \times 2$ matrix

$$[\,\mathbf{p}_\mu\ \mathbf{p}_\sigma\,]$$

where $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$ are the vectors of derivatives of the probabilities $\mathbf{p}$ with respect to μ and σ respectively. The tangent subspace, $\mathfrak{R}(\mathbf{X})$, to the estimation locus is therefore the vector space spanned by $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$, a two dimensional subspace, $\mathcal{T}_2$, of the sample space. Shortly, it will be convenient to have an orthogonal basis of $\mathcal{T}_2$; this can be generated by Gram-Schmidt orthogonalisation:

$$\mathbf{x}_\mu = \mathbf{p}_\mu$$

$$\mathbf{x}_\sigma = \mathbf{p}_\sigma - \frac{\langle \mathbf{p}_\sigma, \mathbf{x}_\mu \rangle}{\|\mathbf{x}_\mu\|^2}\,\mathbf{x}_\mu$$

where $\langle u,v \rangle$ is the inner product on the sample space, and $\|u\| = \langle u,u \rangle^{1/2}$.

Suppose we now consider a three parameter distribution with parameters μ, σ and τ. The estimation is equivalent, as above, to iterated weighted regression on p_μ, p_σ and p_τ. The tangent subspace, $\mathcal{T}_3$, is the three dimensional subspace of sample space spanned by these three vectors, and an orthogonal basis is given by x_μ, x_σ as above, together with the component of p_τ orthogonal to x_μ and x_σ

$$x_\tau = p_\tau - \frac{\langle p_\tau, x_\sigma \rangle}{\|x_\sigma\|^2} x_\sigma - \frac{\langle p_\tau, x_\mu \rangle}{\|x_\mu\|^2} x_\mu$$

From an inferential point of view, it is useful to consider the regression on p_μ, p_σ and p_τ as a regression on p_μ and p_σ with covariance on p_τ, the two being equivalent. As we have noted, heuristically we would expect the threshold parameter to be determined mainly by the information in the smaller observations which, for this study, are of minor interest. The analysis of covariance implies an analysis conditional on the covariate. By fitting the third parameter we are covariancing out this information which will be associated with p_τ.

Since for a given sample vector f, the fitted probabilities depend upon the tangent subspace $\mathcal{R}(x_\mu, x_\sigma, x_\tau)$, the modification of the fitted probabilities in going from a two parameter fit with tangent subspace $\mathcal{R}(x_\mu, x_\sigma)$ to a three parameter fit, depends upon the orthogonalised vector x_τ. The likeness of two specifications with alternative third parameters τ_1 and τ_2 could be measured by the angle ϕ, given by

$$\cos \phi = \frac{\langle x_{\tau_1}, x_{\tau_2} \rangle}{\|x_{\tau_1}\| \, \|x_{\tau_2}\|} \tag{7.30}$$

between their orthogonalised vectors.

We next consider fitting a censored two parameter lognormal distribution, where in this case, we censor data in the lower tail. For grouped data this is equivalent to combining two or more of the lower tail class intervals and fitting a distribution to the resulting coarser class frequencies. Thus for each class censored, an additional one degree of freedom is lost.

Now we can approximate this procedure by retaining the original m classes, and for each degree of freedom lost by censoring, covariancing the regression on a vector $\mathbf{c} = \boldsymbol{\Sigma}\mathbf{c}^*$ in the sample space. As we successively censor classes from the lower tail of the distribution, a suitable sequence of orthogonal vectors $\mathbf{c}_1^*$, $\mathbf{c}_2^*$, . . . in contrast space will be given by (assuming equal class interval lengths)

$$\mathbf{c}_1^* = (\ -1\ \ 1\ \ 0\ \ 0\ \ 0\ \ \ldots\ \ 0\ )'$$

$$\mathbf{c}_2^* = (\ -\tfrac{1}{2}\ \ -\tfrac{1}{2}\ \ 1\ \ 0\ \ 0\ \ \ldots\ \ 0\ )'$$

$$\mathbf{c}_3^* = (\ -1/3\ \ -1/3\ \ -1/3\ \ 1\ \ 0\ \ \ldots\ \ 0\ )'$$

Suppose we combine the first two classes. Then we have $\mathbf{c}^* = \mathbf{c}_1^*$ above. $\mathbf{c}^*$ is mapped into the sample space by $\boldsymbol{\Sigma}$ giving $\boldsymbol{\Sigma}\mathbf{c}^* = \mathbf{c}$, say. As for $\mathbf{p}_\tau$ above, we can calculate the component of $\mathbf{c}$ orthogonal to $\mathbf{p}_\mu$ and $\mathbf{p}_\sigma$ as

$$\mathbf{x}_c = \mathbf{c} - \frac{\langle \mathbf{c}, \mathbf{x}_\sigma \rangle}{\|\mathbf{x}_\sigma\|^2}\,\mathbf{x}_\sigma - \frac{\langle \mathbf{c}, \mathbf{x}_\mu \rangle}{\|\mathbf{x}_\mu\|^2}\,\mathbf{x}_\mu$$

Equation 7.30 then provides a comparison of the two alternative specifications (three parameter lognormal and censored two parameter lognormal) via consideration of the angle between the two vectors $\mathbf{x}_\tau$ and $\mathbf{x}_c$:

$$\cos\phi = \frac{\langle x_\tau, x_c \rangle}{\| x_\tau \| \, \| x_c \|} \tag{7.31}$$

This represents the partial correlation of p_τ and c eliminating p_μ and p_σ.

7.8.3 Approximations to the three parameter and censored two parameter distributions In demonstrating these ideas with numerical examples it is necessary to be able to make direct comparisons between the different fitted distributions. We are interested in comparisons of variances and information at fixed arbitrary values of the parameters, but for the sake of interest, choose typical values which arise from specific examples in practice.

Calculations which we need to make will in general involve the matrices **P**, **Σ**, **X** and **L**. To compare the three distributions we make approximations to the maximum likelihood solutions for the three parameter and censored two parameter cases, so that the fitted probabilities, **p**, are the same as for the two parameter fit. Thus **P** and **Σ** remain the same for the three distributions.

Let μ_2 and σ_2 be the maximum likelihood estimates for the two parameter lognormal distribution. Then as an approximation to the three parameter distribution, we take the three parameter distribution given by μ_2, σ_2 with τ = 0. The censored two parameter distribution is approximated by the two parameter distribution given by μ_2 and σ_2, retaining the original m classes, and covariancing on the vector **c** as explained above.

Then the vectors of orthogonal derivatives which constitute the columns of the **X** matrix for each case are as follows:

two parameter distribution	x_{μ_2}	x_{σ_2}	
approx. three parameter distribution	x_{μ_2}	x_{σ_2}	x_{τ_0}

approx. censored two parameter distribn x_{μ_2} x_{σ_2} x_c

Denoting the maximum likelihood estimates for the three parameter distribution as μ_3, σ_3, τ_3, and those for the censored two parameter distribution as μ_c, σ_c, the equivalent **X** matrices for these distributions have columns as follows:

three parameter distribution	x_{μ_3}	x_{σ_3}	x_{τ_3}
censored two parameter distribution	x_{μ_c}	x_{σ_c}	

A (non-unique) **L** matrix can be determined once the **X** matrix has been calculated.

Calculations were carried out using Matlab (Moler, 1976) running under the Unix operating system on a DEC Vax 11/750 computer.

<u>Adequacy of the approximations.</u> We can use equation (7.27) for the partition of χ^2 to demonstrate that the approximations to the two distributions are adequate. Table 7.5 gives the subdivisions of χ^2 for both the approximate and the exact models, using the Busselton data.

Table 7.5

Partitions of χ^2 for exact and approximate models

Busselton data

Model	m	k	χ^2_{m-1}	χ^2_k	χ^2_{m-k-1}
two parameter	6	2	4.643	0	4.643
three parameter	6	3	1.120	0	1.120
approximation	6	3	4.643	3.326	1.318
censored two param.	5	2	0.404	0	0.404
approximation	6	3	4.643	3.978	0.665

No value in the table is significant at P = 0.05.

Since both approximate distributions have the same fitted values as the two parameter distribution, the total χ^2 values for all three are the same (=4.643). The exact distributions all show the zero component of χ_k^2 which is the component dependent on deviations from the maximum likelihood estimates of the parameters. It is the third component, on $m-k-1$ degrees of freedom, which is of chief interest however. This component tests the deviations of the data from the model. In both the three parameter and the censored two parameter cases, the value of this component is similar for both the exact and approximate distributions. The ratio of the two values will be distributed as an F variable on 2 and 2 degrees of freedom:

for the three parameter lognormal: 1.18 $\sim F_{2,2}$

for the censored two parameter lognormal: 1.65 $\sim F_{2,2}$

Neither of these values approach significance, and we conclude that the approximations are adequate representations of the exact distributions.

For the three parameter case, this conclusion can be further illustrated by calculating the angle, ϕ, or "correlation" between the two vectors $\mathbf{x}_{\tau_3}$ and $\mathbf{x}_{\tau_0}$, using equation (7.30). We find that

$\cos\phi = 0.965$, using **P** from the three parameter distribution

$\cos\phi = 0.969$, using **P** from the approximate three parameter distribution

i.e.

$\phi = 15.3^\circ$ or 14.3°.

so that the two vectors are very nearly parallel, which confirms our previous conclusion.

7.8.4 <u>Comparison of the distributions via the covariates</u> Given, then, these approximations, we can compare the three parameter and censored two parameter lognormal distributions fitted to the Busselton data by calculating the angle, ϕ, between the two covariates x_{τ_0} and x_c, again using equation (7.30). The result

$$\cos \phi = 0.9835$$

giving ϕ = 10.4° or 0.182 radians, demonstrates that the two covariates are very nearly parallel, and so the two distributions are very similar.

7.8.5 <u>The removal of spurious information, part 1</u> Equation (7.26) partitioned the nonparametric estimate $\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ into the parametric estimate of the contrast $\mathbf{q}^{*\prime}(\mathbf{p}-\mathbf{p}_0)$ plus a component whose expectation can be assumed to be zero if the parametric specification is correct:

$$\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0) = \mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0) + \mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0) \tag{7.32}$$

with a corresponding partition of variances (equation (7.29))

$$\frac{1}{n}\mathbf{q}^{*\prime}\Sigma\,\mathbf{q}^{*} = \frac{1}{n}\mathbf{q}_1^{*\prime}\Sigma\,\mathbf{q}_1^{*} + \frac{1}{n}\mathbf{q}_2^{*\prime}\Sigma\,\mathbf{q}_2^{*} \tag{7.33}$$

We now use these decompositions for each of the two parameter, three parameter and censored two parameter lognormal distributions. As in the previous section, we use an approximation in the latter two cases, so that the total variance is the same for all three cases. Table 7.6 gives, for each model, the (scaled) variances of the components of equation (7.32), using the Busselton data and the linear functional for "excess consumption". (As a comparison, the corresponding variances for the exact (maximum likelihood) three parameter distribution are 6.897, 6.797, 0.100; for the exact censored two parameter distribution the variances are 6.970, 6.818, 0.152.)

Table 7.6

Variances (x n) of components of $q^{*\prime}(f-p_0)$

Busselton data, $q^* = (0\ 0\ 0\ 10\ 30\ 50)'$

Model	component		
	$q^{*\prime}(f-p_0)$	$q_1^{*\prime}(f-p_0)$	$q_2^{*\prime}(f-p_0)$
two parameter	11.816	8.675	3.141
approx. three parameter	11.816	11.648	0.168
approx. censored two param.	11.816	11.659	0.157

We see that for all three models, the variance, $V[q^{*\prime}(f-p_0)]$, of the nonparametric estimate is constant. This is the result of approximating the three parameter and censored two parameter distributions. The information for this non-parametric estimate is given by the reciprocal of the variance, i.e. $n/11.816 = 84.72$. We take this as a "base level" of information.

Estimates derived from parametric specifications will have decreased variance, or equivalently, increased information, induced by the assumptions implied by the specification. For the three specifications, this information is given by the reciprocal of the variance of the first component, $q_1^{*\prime}(f-p_0)$, namely

two parameter	115.39
three parameter	85.94
censored two parameter	85.86

The increase in information for the two parameter model is markedly greater than that for either of the other two models, where the information is very close to that for the nonparametric estimate. It is this increase in information which we claim is "spurious" information induced by unwarranted assumptions in the specification, that is, the assumption of strict symmetry on the log scale. Thus both the three parameter and censored two parameter models appear to have substantially eliminated the spurious information.

It is of interest to compare this with the estimation of a contrast, mean consumption, which is not so dependent on the upper tail. The corresponding figures for the information are

two parameter	5.899
three parameter	5.895
censored two parameter	5.894
nonparametric	5.890

In this case the information for all three specifications is very close to that for the nonparametric estimate; the two parameter lognormal may then be considered a valid specification for this particular inference.

We will return to the removal of spurious information later, but firstly we require a further decomposition of linear functionals.

7.8.6 Further decompositions of linear functionals Recall that the decomposition theorem proved earlier gives a means of partitioning contrast space into the direct sum of Σ-orthogonal subspaces

$$\mathscr{C}^* = \mathfrak{R}(P^{-1}X) \oplus \mathfrak{R}(P^{-1}X)^{\perp} \tag{7.34}$$

with dimensions respectively k and $m-k-1$. Equivalently, sample space is partitioned into the direct sum Σ^--orthogonal subspaces

$$\mathscr{S} = \mathfrak{R}(X) \oplus \mathfrak{R}(X)^{\perp} .$$

For the moment, we confine our attention to the contrast space.

The score-functional subspace, $\mathfrak{R}(P^{-1}X)$, is spanned by the score-functionals

$$P^{-1} p_{\theta_i} \qquad i = 1, \ldots, k$$

which are the columns of $P^{-1}X$. Thus a basis of this subspace is given by the Gram-Schmidt orthogonalised vectors

$$P^{-1} x_{\theta_i} \qquad i = 1, \ldots, k.$$

In the case of the three parameter and approximate censored two parameter lognormal models we gave formulae for x_μ, x_σ, x_τ, and x_μ, x_σ, x_c in Section 7.8.2 above.

Thus we can further decompose the score functional subspace into Σ-orthogonal components, one associated with each parameter of the model:

$$\mathfrak{R}(P^{-1}X) = \mathfrak{R}(P^{-1}x_{\theta_1}) \oplus \mathfrak{R}(P^{-1}x_{\theta_2}) \oplus \ldots \oplus \mathfrak{R}(P^{-1}x_{\theta_k})$$

In the case of the three parameter lognormal distribution:

$$\mathfrak{R}(P^{-1}X) = \mathfrak{R}(P^{-1}x_\mu) \oplus \mathfrak{R}(P^{-1}x_\sigma) \oplus \mathfrak{R}(P^{-1}x_\tau)$$

Since our interest does not lie in μ or σ individually, we write $X_{\mu\sigma} = [x_\mu \; x_\sigma]$, and have

$$\mathfrak{R}(P^{-1}X) = \mathfrak{R}(P^{-1}X_{\mu\sigma}) \oplus \mathfrak{R}(P^{-1}x_\tau)$$

Then the contrast space decomposition (7.34) can now be written, in an obvious notation,

$$\mathfrak{C}^* = \mathfrak{R}(P^{-1}X_{\mu\sigma}) \oplus \mathfrak{R}(P^{-1}x_\tau) \oplus \mathfrak{R}(P^{-1}X_{\mu\sigma\tau})^\perp \qquad (7.35)$$

and the previous decomposition of the linear functional q^*

$$(I_m - 1_m p')\, q^* = q_1^* + q_2^*$$

becomes

$$(I_m - 1_m p')\, q^* = q_{1\mu\sigma}^* + q_{1\tau}^* + q_2^* \qquad (7.36)$$

where

$$q^*_{1\mu\sigma} = E^*_{1\mu\sigma}\, q^*$$
$$= P^{-1}X_{\mu\sigma}(X_{\mu\sigma}'P^{-1}X_{\mu\sigma})^{-1}X_{\mu\sigma}'\, q^*$$

and

$$q^*_{1\tau} = E^*_{1\tau}\, q^*$$
$$= P^{-1}x_{\tau}(x_{\tau}'P^{-1}x_{\tau})^{-1}x_{\tau}'\, q^*$$

We can use equation (7.36) to consider further partitions of the non-parametric estimator $q^*{}'(f - p_0)$, giving

$$q^*{}'(f - p_0) = q^*_{1\mu\sigma}{}'(f - p_0) + q^*_{1\tau}{}'(f - p_0) + q^*_2{}'(f - p_0) \tag{7.37}$$

with an equivalent partition of the variance

$$\frac{1}{n}q^*{}'\Sigma\, q^* = \frac{1}{n}q^*{}'\Sigma_{1\mu\sigma}\, q^* + \frac{1}{n}q^*{}'\Sigma_{1\tau}\, q^* + \frac{1}{n}q^*{}'\Sigma_2\, q^* \tag{7.38}$$

where

$$\Sigma_{1\mu\sigma} = X_{\mu\sigma}(X_{\mu\sigma}'P^{-1}X_{\mu\sigma})^{-1}X_{\mu\sigma}'$$
$$\Sigma_{1\tau} = x_{\tau}(x_{\tau}'P^{-1}x_{\tau})^{-1}x_{\tau}'$$

7.8.7 <u>The removal of spurious information, part 2</u> We are now in a position to examine further the apparent removal of information by the three parameter and censored two parameter lognormal distributions demonstrated above using the estimation of "excess consumption" for the Busselton data. We do so by considering the decomposition (7.37) of the contrast $q^*{}'(f - p_0)$.

In fitting a three parameter model, the fit will make the first two terms of the right hand side of (7.37) equal to zero, and under reasonable assumptions, the final term will have expectation zero. However, had we postulated a two parameter model, then only the first term will be made zero by the fit, and the last two terms will be combined. If we accept the implied assumptions of the two parameter model, we are then accepting that the expectation of this combined term is zero. That is, we are also accepting that

$$E[q^{*\prime}_{1\tau}(f - p_0)] = 0.$$

We believe that this is a very strong and questionable assumption for the case of inferences about the upper tail of the distribution of alcohol consumption.

As before, the components of variance given in (7.38) will show the magnitude of the importance of the estimates in (7.37).

These decompositions are also applicable to the censored two parameter distribution. If we use the matrix of derivatives,

$$X_{\mu\sigma c} = [\, x_\mu \; x_\sigma \; x_c \,]$$

for the approximate censored two parameter distribution, then the parallel is obvious.

Table 7.7 gives details of these variance partitions for both models, again using the Busselton data and the "excess consumption" linear functional.

Table 7.7

Further partitions of variances (x n) of $q^{*\prime}(f-p_0)$

Busselton data, $q^* = (0\ 0\ 0\ 10\ 30\ 50)'$

Model	component			
	$q^{*\prime}(f-p_0)$	$q^{*\prime}_{1\mu\sigma}(f-p_0)$	$q^{*\prime}_{1x}(f-p_0)$	$q^{*\prime}_{2}(f-p_0)$
two parameter	11.816	8.675	–	3.141
approx. 3 param. ($x=\tau$)	11.816	8.675	2.973	0.168
approx. cens. 2 param. ($x=c$)	11.816	8.675	2.984	0.157

In the previous section, we saw that for both the three parameter and censored two parameter models, the variance associated with the parametric estimate $q^{*\prime}(\hat{p} - p_0)$ was increased, or equivalently, the information reduced, in comparison with the same contrast from the two parameter model. Table 7.7 shows clearly that this increase of variance is associated with the covariates x_τ and x_c introduced respectively by the third parameter τ, and by censoring the two parameter distribution. Since for both these distributions the variance associated with the covariate is $R^2 = 25\%$ of the total variance, we would be unwilling to accept the expectation of this term as zero, as would be demanded by a two parameter specification. This is despite the fact that that a χ^2 goodness-of-fit test for the two parameter distribution gives a non-significant result (see Table 7.5).

This value of R^2 contrasts markedly with that obtained using the functional q_t^* (see Table 7.2) to estimate "mean consumption". In that case, the variance associated with the covariate is, for both modified distributions, less than 0.1% of the variance of the nonparametric estimate. The variance of the estimate of mean consumption derived from the two parameter distribution accounts for more than 98.8% of the variance of the nonparametric estimate, and we would have little hesitation in concluding that, for estimating this particular contrast, the two parameter distribution was valid.

Further information is provided by the elements, and the norm or length of the vector $q_{1\tau}^*$ (or q_{1c}^*). This vector is the projection of q^* on $\mathcal{R}(P^{-1}x_\tau)$:

$$q_{1\tau}^* = E_{1\tau}^* q^*$$

Since $E_{1\tau}^*$ is a projection matrix which depends only on the specification and not on q^*, for a given specification the projections of two different q^*

vectors will be parallel but have different norms. Suppose that, for a given specification, ${}_{a}\mathbf{q}^{*}_{1\tau}$ and ${}_{b}\mathbf{q}^{*}_{1\tau}$ are the projections on $\mathfrak{R}(\mathbf{P}^{-1}\mathbf{x}_{\tau})$ of the two linear functionals $\mathbf{q}^{*}_{a}$ and $\mathbf{q}^{*}_{b}$. Then since the two are parallel, their elements will be proportional

$${}_{b}\mathbf{q}^{*}_{1\tau} = k\ {}_{a}\mathbf{q}^{*}_{1\tau}$$

for some constant k. Thus in considering the elements of $\mathbf{q}^{*}_{1\tau}$, scale is unimportant, and the relative sizes of the elements of $\mathbf{q}^{*}_{1\tau}$ will indicate the nature of the information the vector highlights; this will be the same for *all* $\mathbf{q}^{*}$ vectors, given a particular specification.

The magnitude of the norm of $\mathbf{q}^{*}_{1\tau}$, in relation to the norm of $\mathbf{q}^{*}_{1\mu\sigma}$, will indicate the importance of the inclusion in the specification of the covariate $\mathbf{x}_{\tau}$ (or $\mathbf{x}_{c}$), and this will be different for different $\mathbf{q}^{*}$ vectors.

Table 7.8 gives the vectors $\mathbf{q}^{*}_{1\tau}$ for the three parameter distribution and the approximations to the three parameter and censored two parameter distributions, fitted to the Busselton data, scaled for convenience so that the largest element is unity.

Table 7.8

Scaled vectors $\mathbf{q}^{*}_{1\tau}$ for three lognormal specifications

Busselton data

3 param.	approx. 3 param.	approx. cens. 2 param.
-.002	-.002	-.003
.050	.071	.084
-.182	-.335	-.432
-.082	-.182	-.084
.241	.138	.276
1.000	1.000	1.000

From these vectors it is clear that $q^*_{1\tau}$ is very heavily weighted to those frequencies in the upper tail. While generalisations from examinations of particular data sets in such detail must be seasoned with caution, a general interpretation is that $q^*_{1\tau}$ contrasts the upper tail frequencies against those in the middle regions and lower tail of the distribution. Random fluctuations in the data disturb this pattern only to a minor extent.

Thus, irrespective of which q^* linear functional we consider, the addition of the extra parameter τ, or the censoring of the lower tail, serves to weight the upper tail frequencies, and the estimation no longer rests on the symmetry assumption implicit in the two parameter lognormal.

It is in considering the estimation of a particular contrast, that is, using a particular linear functional q^*, that the norm or length of the projection is important. The norms are given by

$$\|q^*_{1\mu\sigma}\| = (q^{*\prime}_{1\mu\sigma} \Sigma \, q^*_{1\mu\sigma})^{1/2}$$

and

$$\|q^*_{1\tau}\| = (q^{*\prime}_{1\tau} \Sigma \, q^*_{1\tau})^{1/2}$$

Table 7.9 gives the norms of $q^*_{1\mu\sigma}$ and $q^*_{1\tau}$ (or q^*_{1c}) for the linear functionals q^*_e (excess consumption) and q^*_t (mean consumption) for the three lognormal distributions. Also given is the ratio of the two norms, which is the length of $q^*_{1\tau}$ if the projections are normed so that $q^*_{1\mu\sigma}$ has unit length. We may note several points from the table.

Firstly it confirms that the approximation to the three parameter distribution is a reasonable one, as results for the approximate distribution agree closely with those for the true maximum likelihood distribution. This was

Table 7.9

Norms of projections of q* on $\Re(P^{-1}X_{\mu\sigma})$ and $\Re(P^{-1}x_{\tau})$

Busselton data

q*	model	$\|q^*_{1\mu\sigma}\|$	$\|q^*_{1\tau}\|$	ratio
q^*_e	3 parameter	2.212	1.379	.62
	approx 3 parameter	2.945	1.724	.58
	approx censored 2 param	2.945	.728	.25
q^*_t	3 parameter	12.650	.084	.01
	approx 3 parameter	13.027	.330	.03
	approx censored 2 param	13.027	.382	.03

also confirmed by the previous table.

Secondly, it again demonstrates the effect of the approximations in facilitating comparisons between the distributions, as, for a given linear function q*, both approximated distributions have equal length projections on $\Re(P^{-1}X_{\mu\sigma})$. This is also the norm of the similar projection for the two parameter lognormal.

Finally, and most importantly, is a consideration of the ratios

$$\frac{\|q^*_{1\tau}\|}{\|q^*_{1\mu\sigma}\|}$$

For the excess consumption functional, q^*_e, both the three parameter and censored two parameter distributions have a substantial projection on $\Re(P^{-1}x_{\tau})$, while for the mean consumption functional, q^*_t, the norms of the projections, which are associated with the covariates x_τ or x_c, are close to zero. Again this demonstrates that while the two parameter distribution is a valid distribution for estimating $q^{*\prime}_t p$, the extra parameter, or censoring the lower tail, are necessary to ensure validity for estimation of contrasts such as $q^{*\prime}_e p$ which are concerned with the upper tail.

7.8.8 Discussion and summary Alcohol researchers since Ledermann's first attempt in 1956 have attempted to find a single specification to describe the distribution of individual alcohol consumption, and for use in making inferences about the distribution.

Such a specification must necessarily be as simple as possible. As we have seen in Part I of this thesis, alcohol consumption data is rarely available with more than six frequency classes. This means that often there is simply not enough data to attempt to fit more complex specifications, such as the log-hypergeometric, which may otherwise have more desirable features.

Thus one of the commonly used specifications for the distribution of alcohol consumption has been the two parameter lognormal. If after fitting this distribution, a three parameter lognormal was fitted, but it was found that the third parameter was not significantly different from zero (suppose it was less than its standard error) then it may appear at first sight that the two parameter fit was preferable on the grounds of simplicity.

But in choosing a specification, there are important considerations beyond the goodness-of-fit, namely the reliance on questionable assumptions. Some of the assumptions underlying a specification may be reasonable, e.g. that the distribution is "smooth". Others may have neither inherently compelling reasons nor factual foundation, but involve mere assumptions perhaps adopted for convenience. Inferences depending heavily upon unjustified assumptions will be suspect, and in some cases even mischievous. These ideas are embodied in the relevance principle, as given by Wilkinson (1977) after Fisher (1973). This principle requires that inferences involving uncertainty should use all relevant information, both quantitative and qualitative, and exclude irrelevant or spurious information.

The two parameter lognormal distribution provides an example of assumptions which may be unjustified. The distribution assumes a strict symmetry on the log scale. In making inferences about the upper tail, we may not wish to rely heavily upon this assumption, even though the data do not give significant evidence of it being violated. In the alcohol case, information about the light drinkers may be being spuriously used to make inferences about the heavy drinkers.

In exploring this situation we showed that the nonparametric estimate of a contrast could be partitioned as

$$\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0) = \mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0) + \mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$$

into the parametric estimate plus an estimator of zero. In using the parametric estimate, we are depending on the specification to assume that the second component has zero expectation.

The symmetry assumption of the two parameter lognormal appears to invalidate the zero expectation of the second component for inferences concerning the upper tail. We have shown that for the estimation of a contrast concerned largely with the centre of the distribution (the mean consumption), the two parameter specification was adequate, whereas for a contrast concerned largely with the upper tail (excess consumption), a substantial proportion of the variance of the nonparametric estimate $\mathbf{q}^{*\prime}\mathbf{f}$ was not accounted for by the parametric estimate $\mathbf{q}^{*\prime}\hat{\mathbf{p}}$ of the contrast. This is an example of the noncoherence principle (Wilkinson, 1977), in that the specification is valid for one inference, but invalid for another.

We have suggested two possible modifications of the two parameter lognormal to improve the situation, namely, adding a third parameter, or censoring the lower tail of the distribution. Using a covariance approach to fitting these distributions, we demonstrated their similarity.

To see the effect of the assumption of symmetry we demonstrated the quantitative effects of removing it. In the estimation of excess consumption, a contrast concerned particularly with the upper tail, either adding the third parameter or censoring the distribution gave variances of the resulting parametric estimates which were much closer to the variance of the non-parametric estimate than was the variance of the two parameter estimate. This increase in variance for the estimates from the modified distributions corresponds to a decrease in information, as compared to the two parameter lognormal, and we claim that the information so lost is spurious, being based on the unwarranted assumption of strict symmetry of the distribution on the log scale.

The third parameter, τ, when added to the two parameter lognormal, produces a covariate vector, $\mathbf{p}_\tau$, which transfers a component from $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ to $\mathbf{q}_1^{*\prime}(\mathbf{f}-\mathbf{p}_0)$. The covariate vector $\mathbf{c}$ introduced by censoring the two parameter distribution has a similar effect. The component so transferred is heavily weighted to frequencies in the upper tail, freeing the estimation from the symmetry assumption implicit in the two parameter lognormal distribution. Thus using either of these modified specifications, we are more confident that the expectation of the new reduced $\mathbf{q}_2^{*\prime}(\mathbf{f}-\mathbf{p}_0)$ is zero.

Thus, for inferences concerning the upper tail of the distribution of alcohol consumption, we have greater confidence in the validity of either the three parameter lognormal or the censored two parameter lognormal distribution than we do in the two parameter lognormal distribution.

7.9 Fitting a distribution subject to a constraint on a linear function of the fitted probabilities.

7.9.1 Introduction This section is not directly associated with the previous sections of this chapter, but gives a related method of estimation of contrasts. While the previous sections have been largely concerned with the lognormal distribution, the theory given here does not depend on any particular distribution.

We have been concerned with partitions of contrasts, which we can write as

$$\mathbf{q}^{*\prime}(\mathbf{f}-\mathbf{p}_0) = \mathbf{q}^{*\prime}(\hat{\mathbf{p}}-\mathbf{p}_0) + \mathbf{q}^{*\prime}(\mathbf{f}-\hat{\mathbf{p}}) \; .$$

As we have said, this partitions the nonparametric estimate of $\mathbf{q}^{*\prime}(\mathbf{p}-\mathbf{p}_0)$ into two components, the parametric estimate $\mathbf{q}^{*\prime}(\hat{\mathbf{p}}-\mathbf{p}_0)$, plus a component $\mathbf{q}^{*\prime}(\mathbf{f}-\hat{\mathbf{p}})$ whose expectation is zero on the assumption that the parametric specification is correct.

This leads us to consider fitting the parametric distribution constraining the component $\mathbf{q}^{*\prime}(\mathbf{f}-\hat{\mathbf{p}})$ to be equal to zero.

Suppose from survey results we were interested in estimating the amount of alcohol consumed in excess of 100 g. per day. It may be known that greater effort has been expended on interviews with respondents who reported alcohol consumption greater than 60 g. per day (the "safe" limit, according to some medical authorities), so there is reason to believe that data above 60 g./day is more accurate than that in the rest of the distribution. Since the required inference concerns the extreme upper tail of the distribution, it is reasonable that we may wish to place greater weight on the data in that area than on the rest of the data.

In such a situation, we suggest that it may be appropriate to base the inference on a distribution which has been fitted subject to the constraint that

$$\mathbf{q}^{*\prime}\hat{\mathbf{p}} = \mathbf{q}^{*\prime}\mathbf{f}$$

where in this case, $\mathbf{q}^*$ represents the excess consumption above 60 g. per day, that is, the linear functional $\mathbf{q}^*_e$ of Table 7.2.

The mathematics to achieve this is straightforward, using the iterated regression formulation of maximum likelihood estimation given in Section 7.4.3, and imposing the linear constraint by Lagrange multipliers. We give this in the following section.

7.9.2 <u>Fitting the model</u> Section 7.4.3 showed that maximum likelihood estimation can be regarded as iterative weighted regression of $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{f} - \mathbf{p}$ on $\mathbf{X}$, the matrix of derivatives of the class probabilities with respect to the parameters, with $\mathbf{P}^{-1}$ as weight matrix. The residual sum of squares from this regression is then

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'\mathbf{P}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$$

Our approach is to minimise this subject to the linear constraint

$$\mathbf{q}^{*\prime}(\mathbf{f} - \hat{\mathbf{p}}) = 0\ .$$

Suppose we start the iterative process at $\boldsymbol{\theta} = \boldsymbol{\theta}_r$. Then we have

$$\mathbf{y}_r = \mathbf{X}_r\boldsymbol{\theta}_r + \mathbf{f} - \mathbf{p}_r \qquad (7.39)$$

and the residual sum of squares is

$$(\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\theta}_r)'\mathbf{P}_r^{-1}(\mathbf{y}_r - \mathbf{X}_r\boldsymbol{\theta}_r) \qquad (7.40)$$

and the constraint is

$$\mathbf{q}^{*\prime}(\mathbf{f} - \mathbf{p}_r) = 0\ .$$

Substituting for $\mathbf{f} - \mathbf{p}_r$ from (7.39), the constraint becomes

$$q^{*\prime}(y_r - X_r\theta_r) = 0 \ . \tag{7.41}$$

We use a Lagrange multiplier, 2λ, to achieve the minimisation, and from (7.40) and (7.41) write the residual sum of squares and the constraint as

$$(y_r - X_r\theta_r)' P_r^{-1} (y_r - X_r\theta_r) + 2\lambda q^{*\prime}(y_r - X_r\theta_r) \tag{7.42}$$

Differentiating with respect to θ_r and equating to zero gives

$$-2\, d\theta_r' X_r' P_r^{-1}(y_r - X_r\theta_r) - 2\lambda d\theta_r' X_r' q^* = 0$$

Equating coefficients of $2d\theta_r'$:

$$X_r' P_r^{-1} (y_r - X_r\theta_r) = \lambda X_r' q^*$$

Expanding and rearranging gives the basis of the iterative scheme

$$\theta_{r+1} = (X_r' P_r^{-1} X_r)^{-1} X_r' P_r^{-1} y_r - \lambda (X_r' P_r^{-1} X_r)^{-1} X_r' q^* \tag{7.43}$$

Now substituting this value for θ into (7.41) gives, approximately

$$q^{*\prime}(y_r - X_r[(X_r' P_r^{-1} X_r)^{-1} X_r' P_r^{-1} y_r - \lambda (X_r' P_r^{-1} X_r)^{-1} X_r' q^*]) = 0$$

Rearranging this equation gives

$$\lambda = \frac{q^{*\prime} X_r (X_r' P_r^{-1} X_r)^{-1} X_r' P_r^{-1} y_r - q^{*\prime} y_r}{q^{*\prime} X_r (X_r' P_r^{-1} X_r)^{-1} X_r' q^*}$$

$$= \frac{s_1}{s_2} \quad \text{say}$$

Thus from (7.43)

$$\theta_{r+1} = (X_r' P_r^{-1} X_r)^{-1} X_r' P_r^{-1} y_r - \frac{s_1}{s_2} (X_r' P_r^{-1} X_r)^{-1} X_r' q^*$$

which on substituting for y_r from (7.39) yields

$$\theta_{r+1} = \theta_r + (X_r' P_r^{-1} X_r)^{-1} X_r' P_r^{-1} (f - p_r) - \frac{s_1}{s_2} (X_r' P_r^{-1} X_r)^{-1} X_r' q^* \tag{7.44}$$

where

$$s_1 = q^{*\prime} X_r (X_r' P_r^{-1} X_r)^{-1} X_r P_r^{-1} y_r - q^{*\prime} y_r$$

which we see, using equations (7.39) and (7.11)

$$= -q^{*\prime}(f - \hat{p})$$

and

$$s_2 = q^{*\prime} X_r (X_r' P_r^{-1} X_r)^{-1} X_r' q^*$$
$$= V[\, q^{*\prime} \hat{p}\,]$$

We can now iterate (7.44) until θ is arbitrarily accurate.

7.9.3 Example Table 7.10 gives a comparison of the maximum likelihood two parameter lognormal distribution fitted to the Busselton data, and the fit subject to the constraint

$$q^{*\prime}(f - \hat{p}) = 0$$

where $q^* = (0\ 0\ 0\ 10\ 30\ 50)'$, ("excess consumption").

Table 7.10

Comparison of maximum likelihood and constrained fits

Busselton data, two parameter lognormal, $q^*=(0\ 0\ 0\ 10\ 30\ 50)'$

class int.	f	max. likelihood $\hat{p}$	χ^2	constrained fit $\hat{p}$	χ^2	q^*
0–20	.8032	.8020	.002	.8054	.006	0
20–40	.1419	.1497	.408	.1535	.878	0
40–60	.0420	.0320	3.105	.0289	5.859	0
60–80	.0090	.0096	.042	.0077	.199	10
80–100	.0020	.0036	.689	.0026	.135	30
>100	.0020	.0031	.398	.0019	.006	50
			4.643		7.083	

Starting with the maximum likelihood estimates of the two parameter distribution, the fit converged to its final solution rapidly. Parameter estimates were determined to 4 decimal places after 5 iterations, and to 10 decimal places after 10 iterations. The parameter estimates and their standard errors for the two distributions are as follows.

two parameter μ = 2.2717, s.e. = 0.847

σ = 0.8529, s.e. = 0.587

constrained fit μ = 2.3147, s.e. = 0.908

σ = 0.7910, s.e. = 0.623

An examination of the fitted values for the two distributions shows the effect of the constraint. The fitted probabilities for the unconstrained fit considerably overestimate the relative frequencies in the upper tail, where the amount of data available to determine the fit is small. This is despite the fact that the χ^2 value (4.64 on 3 degrees of freedom) is non-significant.

The situation has been improved however for the constrained fit. The agreement between the fitted values and the relative frequencies for the two uppermost classes has been substantially improved. The fit for the 60-80 g/day class has worsened, but the overall fit for these three classes has been improved, as shown by a reduction in the sum of the components of χ^2 for those classes from 1.129 to 0.340.

The fit in the lower tail and the middle regions of the distribution is not as good as for the unconstrained model, but as we have stressed before, in the present application we are not primarily interested in these regions of the distribution, but need to retain them as they contain most of the data.

References.

Advisory Committee on Alcoholism, Department of Health and Social Security and Welsh Office (1977). *Report on Prevention.* London: H.M.S.O.

Aiken, G.J.M. and McCance, C. (1982). Alcohol consumption in offshore oil rig workers. *Brit. J. Addict.* **77**, 305-310.

Aitchison, J. and Brown, J.A.C. (1954). On criteria for descriptions of income distribution. *Metroeconomica* **6**, 88.

Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution with Special Reference to its Uses in Economics.* London: C.U.P.

Armor, D.J. and Polich, J.M. (1982). Measurement of alcohol consumption. Chapter 6 in: **Pattison, E.M. and Kaufman E. (eds.)** *Encyclopedic Handbook of Alcoholism.* New York: Gardner Press 72-80.

Australian Associated Brewers (1978). Methods of measurement - an industry perspective. *Comm. Health Stud.* **2(3)**, 120-122.

Australian Bureau of Statistics. (1974-75). *Apparent Consumption of Foodstuffs and Nutrients.* Catalogue No. 4306.0. Canberra: A.B.S.

Australian Bureau of Statistics. (1978). *Alcohol and Tobacco Consumption Patterns, February 1977.* Catalogue No. 4312.0. Canberra: A.B.S.

Australian Medical Association (1980). Statement on alcohol consumption and abuse. *Med. J. Aust* **2**, 680.

Baghurst, K.I. (1978). A short review of the international literature on the measurement of alcohol consumption and drinking behaviour. *Comm. Health Stud.* **2(3)**, 154–157.

Baghurst, K.I. and Dwyer, T. (1981). Alcohol consumption and blood pressure in a group of young Australian males. *J. Human Nutrition* **35**, 257–264.

Baghurst, K.I. and McMichael, A.J. (1978). Evaluation of questionnaire methods of measurement of alcohol consumption in young Australians. *Comm. Health Stud.* **2(3)**, 135–139.

Bailey, N.J.T. (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford: O.U.P.

Barwon Regional Association for Alcohol and Drug Dependence. (1977). *Alcohol and Drug Use – the Geelong Area*. Geelong: B.R.A.A.D.D.

Bell, D.S., Champion, R.A. and Rowe, A.J.E. (1975). *Monitoring alcohol use among young people in N.S.W. 1971–1973*. Report 75/22, Division of Health Services Research, Health Commission of N.S.W.

Bingham, C. (1980). Distribution on the sphere. In: *R.A. Fisher: An Appreciation*, S.E. Fienberg and D.V. Hinkley (eds) New York: Springer-Verlag. 171–181.

Blair, E., Sudman, S., Bradburn, N.M. and Stocking, C. (1977). How to ask questions about drinking and sex: response effects in measuring consumer behaviour. *J. Marketing Res.* **14**, 316–321.

Bliss, C.I. (1935). The calculation of the dosage-mortality curve. Appendix by R.A. Fisher. *Ann. Appl. Biol.* **22**, 134–167.

Bliss, C.I. (1937). The calculation of the time-mortality curve. Appendix: The truncated normal distribution (by W.L. Stevens) pp 847-850. *Annals Appl. Biol.* **24**, 815-852.

Bliss, C.I. (1967). *Statistics in Biology. Volume 1.* New York: McGraw-Hill Book Co.

Borkenstein, R.F., Crowther, R.F., Shumate, R.P., Ziel, W.B., and Zylman, R. (Dale, A.B. ed.) (1964). *The role of the drinking driver in traffic accidents.* Bloomington: Indiana University, Department of Police Administration.

Bowman, R.S., Stein, L.I. and Newton, J.R. (1975). Measurement and interpretation of drinking behaviour. I. On measuring patterns of alcohol consumption. II. Relationships between drinking behaviour and social adjustment in a sample of problem drinkers. *J. Stud. Alc.* **36**, 1154-1172.

Brezard, M. (1958). Presentation d'une enquête sur la consommation des boissons en France. *Bull. de l'I.N.H.* **13**, 267-358.

Brezard, M. (1959). La consommation des boissons en France. Deuxième partie: Marseille. *Bull. de l'I.N.H.* **14**, 95-163.

Brezard, M. (1960). La consommation des boissons en France. Quatrième partie: quelques districts ruraux. *Bull. de l'I.N.H.* **15**, 229-263.

Brun-Gulbrandsen, S. (1976). *Alkoholbruk blaut norsk ungdom.* Oslo: Universitetsforlaget.

Bruun, K., Edwards, G., Lumio, M., Mäkelä, K., Pan, L., Popham, R.E., Room, R., Schmidt, W., Skog, O.-J., Sulkunen, P., and Österberg, E.

(1975). *Alcohol Control Policies in Public Health Perspective*. Helsinki: The Finnish Foundation for Alcohol Studies, Volume 25.

Cahalan, D., Cisin, I.H. and Crossley, H.M. (1967). *American Drinking Practices: A National Survey of Behaviour and Attitudes Related to Alcoholic Beverages*. Report No. 3, The George Washington University Social Research Group, Washington, D.C.

Cahalan, D., Cisin, I.H. and Crossley, H.M. (1969). *American Drinking Practices: A National Study of Drinking Behaviour and Attitudes*. Monograph No. 6, Rutgers Center of Alcohol Studies, New Brunswick, N.J.

Carrington-Smith, D. (1978). Do women need drugs to cope with life? In: *Women's Health in a Changing Society*. Canberra: A.G.P.S. 240-243.

Cartwright, A. (1977). Population surveys and the curve. In: *The Ledermann Curve: Report of a Symposium held in London on 6 and 7 January, 1977*. London: Alcohol Education Centre. 58-69.

Cartwright, A., Shaw, S.J. and Spratley, T.A. (1977). The validity of *per capita* alcohol consumption as an indicator of the prevalence of alcohol related problems: An evaluation based on national statistics and survey data. In: **J.S. Madden, R. Walker and W.H. Kenyon. (eds.)** *Alcoholism and Drug Dependence. A Multidisciplinary Approach*. London: Plenum Press.

Cartwright, A.K.J, Shaw, S.J. and Spratley, T.A. (1978a). The relationships between *per capita* consumption drinking patterns and alcohol related problems in a population sample 1965-1974. I. Increased consumption and changes in drinking patterns *Brit. J. Addict.* **73,** 237-246 .

Cartwright, A.K.J., Shaw, S.J. and Spratley, T.A. (1978b). The relationship between *per capita* consumption drinking patterns and alcohol related problems in a population sample 1965–1974. II. Implications for alcohol control policy. *Brit. J. Addict.* **73,** 247–258.

Castro, E., Chao, Z. and Smart, R.G. (1978). The distribution of drug use in Mexico: data from a national survey. *Bull. on Narcotics* **30(2),** 49–54.

Chick, J., Kreitman, N. and Plant, N. (1981). Saving face? Survey respondents who claim their last week's drinking was atypical. *Drug and Alc. Depend.* **7,** 265–272.

Cohen, A.C. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *J. Amer. Statist. Assoc.* **46,** 206–212.

Cooke, K.M., Frost, G.W., Thornell, I.R. and Stokes, G.S. (1982). Alcohol consumption and blood pressure. Survey of the relationship at a health-screening clinic. *Med. J. Aust.* **1,** 65–69 (correction p 209; correspondence pp 481–482).

Cox, D.R. (1961). Tests of separate families of hypotheses. *Proc. 4th Berkeley Symposium* **1,** 105–123.

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B.* **24,** 406–424.

Cox, D.R. and McCullagh, P. (1982). Some aspects of analysis of covariance. *Biometrics* **38(3),** 541–561.

Cozby, P.C. (1973). Self-disclosure: A literature review. *Pychological Bulletin* **79,** 73–91.

Cullen, K.J., Stenhouse, N.S., McCall, M.G., Wearne, K.L. and Murphy, B.P. (1980). Alcohol consumption and cigarette smoking in Busselton, 1966-1978. *Med. J. Aust.* **2,** 87-92.

Cullen, K.J. and Woodings, T. (1975). Alcohol, tobacco and analgesics - Busselton 1972. *Med. J. Aust.* **2,** 211-214.

Curnow, D.H., Cullen, K.J., McCall, M.G., Stenhouse, N.S., and Welborn, T.A. (1969). Health and disease in a rural community. A W.A. study. *Aust. J. Sci.* **31(8),** 281-285.

De Lint, J. (1974). The epidemiology of alcoholism: The elusive nature of the problem, estimating the prevalence of excessive alcohol use and alcohol-related mortality, current trends and the issue of prevention. In: **Kessel *et al* (ed.)** *Alcoholism: A Medical Profile; Proc. First International Medical Conference on Alcoholism.* London: Edsall.

De Lint, J. (1976). Review of *Alcohol control policies in public health perspective* by Bruun *et al.* *J. Stud. Alc.* **37,** 1497-1499.

De Lint, J. (1977). The frequency distribution of alcohol consumption: an overview. In: *The Ledermann Curve. Report of a Symposium held in London on 6 and 7 January 1977.* London: The Alcohol Education Centre, 1-10.

De Lint, J. (1978). Total alcohol consumption and rates of excessive use: a reply to Duffy and Cohen. *Brit. J. Addict.* **73,** 265-269.

De Lint, J. and Schmidt, W. (1968). The distribution of alcohol consumption in Ontario. *Quart. J. Stud. Alc.* **29,** 968-973.

De Lint, J. and Schmidt, W. (1971a). Alcohol use and alcoholism. *Addictions* **18(2),** 1–15.

De Lint, J. and Schmidt, W. (1971b). Consumption averages and alcoholism prevalence: a brief review of epidemiological investigations. *Brit. J. Addict.* **66,** 97–107.

Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Reading, Mass.: Addison-Wesley Publishing Co.

Dight, S. (1976). *Scottish Drinking Habits.* London: Office of Population Censuses and Surveys, Social Survey Division, H.M.S.O.

Duffy, J.C. (1977a). Estimating the proportion of heavy consumers. In: *The Ledermann Curve. Report of a Symposium held in London on 6 and 7 Janurary 1977.* London: The Alcohol Education Centre, 11–24.

Duffy, J.C. (1977b). Alcohol consumption, alcoholism and excessive drinking – errors in estimates from consumption figures. *Int. J. Epidemiology.* **6(4),** 375–379.

Duffy, J.C. (1980). The association between *per capita* consumption of alcohol and the proportion of excessive consumers – a reply to Skog. *Brit. J. Addict.* **75,** 147–151.

Duffy, J.C. (1982). Fallacy of the distribution of alcohol consumption. *Psychological Reports* **50,** 125–126.

Duffy, J.C. and Cohen, G.R. (1978). Total alcohol consumption and excessive drinking. *Brit. J. Addict.* **73,** 259–264.

Edwards, G. (1973). Epidemdiology applied to alcoholism: a review and examination of the purposes. *Quart. J. Stud. Alc.* **34**, 28–56.

Egger, G., Champion, R. and Trebilco, P. (1978). *Adolescent Drug and Alcohol use in N.S.W. 1971–1977*. Report no. 78/3, Division of Health Services Research, Health Commission of N.S.W.

Egger, G., Parker, R. and Trebilco, P (1976). *Adolescents and Alcohol in N.S.W.: A Report to the Child Health Committee of the N.S.W. Health Education Advisory Council*. N.S.W., Govt. Printer.

Eichhorn, B.H. and Hayre, L.S. (1983). Scrambled randomised response methods for obtaining sensitive quantitative data. *J. Statist. Planning and Inference* **7**, 307–316.

Ekholm, A. (1972). The lognormal distribution of blood alcohol concentrations in drivers. *Quart. J. Stud. Alc.* **33**, 508–512.

Encel, S. and Kotowicz, K. (1970). Heavy drinking and alcoholism: a preliminary report. *Med. J. Aust.* **1(12)**, 607–612.

Encel, S., Kotowicz, K.C. and Resler, H.E. (1972). Drinking patterns in Sydney, Australia. *Quart. J. Stud. Alc.* **6**, 1–27.

Engs, R.C. (1982). Drinking patterns and attitudes towards alcoholism of Australian human-service students. *J. Stud. Alc.* **43(5)**, 517–531.

Finney, D.J. (1952). *Statistical Methods in Biological Assay*. New York: Hafner Press.

Fisher, R.A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. Roy. Statist. Soc.* **85**, 87–94.

Fisher, R.A. (1931). The sampling error of estimated deviates, together with other illustrations of the properties and applications of the integrals and derivatives of the normal error function. *Mathematical Tables* **1**, xxxvi–xxxv, British Association for the Advancement of Science.

Fisher, R.A. (1935). The detection of linkage with "dominant" abnormalities. *Ann. Eugenics* **6**, 187–201.

Fisher, R.A. (1950). The significance of deviations from expectation in a Poisson series. *Biometrics* **6**, 17–24.

Fisher, R.A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. London* **A 217**, 295–305.

Fisher, R.A. (1954). The analysis of variance with various binomial transformations. *Biometrics* **10**, 130–139.

Fisher, R.A. (1963). *Statistical Methods for Research Workers (13th edition).* Edinburgh: Oliver and Boyd.

Fisher, R.A. (1973). *Statistical Methods and Scientific Inference (3rd edition).* New York: Hafner Press.

Furst, C.J. (1983). Estimating alcoholic prevalence. Chapter 10 in: **Galanter, M. (ed.)** (1983). *Recent Developments in Alcoholism, Vol. 1.* New York: Plenum Press, 269–284.

Gauss, C.F. (1809). *Theoria motus corporum coelestium.* Hamburg.

George, Anne. (1972). Survey of drug use in a Sydney suburb. *Med. J. Aust.* 2, 233-237.

George, Anne. (1973). The survey of drug use in the suburb of Manly, Sydney. In: **Krupinski J. and Stoller A. (eds.)** (1973). *Drug use by the Young Population of Melbourne.* Melbourne: Mental Health Authority.

George Anne. (1974). *1973 Survey of Drug use in a Western Suburb of Sydney.* Unpublished report, N.S.W. Health Commission.

Gibrat, R. (1931). *Les inegalités économiques.* Paris: Libraire de Recueil.

Gibson, J., Johansen, A., Rawson, G. and Webster, I. (1977). Drinking, smoking and drug taking patterns in a predominantly lower socio-economic status sample. Comparison with medicheck sample. *Med. J. Aust.* 2, 459-461.

Gjeddebaek, N.F. (1949). Contributions to the study of grouped observations. Application of the method of maximum likelihood in case of normally distributed observations. *Skand. Aktuartidskr.* 32, 135-157.

Gjeddebaek, N.F. (1956). Contribution to the study of grouped observations. II. Loss of information caused by grouping of normally distributed observations. *Skand. Aktuartidskr.* 39, 154-159.

Graves, G. (1973). Epidemiology of drug use in Melbourne. In: **Krupinski, J. and Stoller, A. (eds.)** (1973). *Drug use in the Young Population of Melbourne.* Melbourne: Mental Health Authority.

Graves, G.D. (1977). *A Survey of drug use in a Rural City.* Melbourne: Govt. Printer.

Graves, G. and Travers, D.J. (1977). *Alcohol and industry*. Melbourne: Institute of Mental Health Research and Postgraduate Training.

Green, P.J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with Discussion). *J. Roy. Statist. Soc. Ser. B* **46**, 149–192.

Grichting, W.L. (1983). Controlling alcohol abuse in Australia: from treatment to prevention. *Brit. J. Addict.* **78**, 37–50.

Grundy, P.M. (1952). The fitting of grouped truncated and grouped censored normal distributions. *Biometrika* **39**, 252–259.

Guttorp, P. and Song, H.H. (1977). A note on the distribution of alcohol consumption. *Drinking and Drug Practices Surveyor* **13**, 7–8.

Guttorp, P. and Song, H.H. (1979). A rejoinder to Skog. *Drinking and Drug Practices Surveyor* **14**, 6 and 29–30.

Hald, A. (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuartidsk.* **32**, 119–134.

Hennesy, B.L., Bruen, W.J. and Cullen, J. (1973). The Canberra Mental Health Survey. *Med. J. Aust.* **1**, 721–728.

Hill, B.M. (1963). The three parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *J. Amer. Statist. Soc.* **58**, 72–84.

Holcomb, R.L. (1938). Alcohol in relation to traffic accidents. *J. Amer. Med. Assoc.* **111**, 1076–1085.

Hyland, J. and Scott, S. (1969). *Alcohol Consumption Tables: An Application of the Ledermann Equation to a Wide Range of Consumption Averages.* Toronto, Canada: Alcohol Research Foundation (mimeograph).

Irwin, R.P. (1976). *Drug education programs and the adolescent in the drug phenomena problem.* Canberra: Department of Sociology, Faculty of Arts, Australian National University.

Jackson. O.A.Y. (1968). Some results of tests of separate families of hypotheses. *Biometrika* **55**, 355-363.

Jackson, O.A.Y. (1969). Fitting a gamma or lognormal distribution to fibre diameter measurements on wool tops. *J. Roy. Statist. Soc. Ser. C Applied Statistics.* **18**, 70-75.

James, A.T. (1973). The variance information manifold and the functions on it. In: **Krishnaiah, P. (ed.)** (1973). *Multivariate Analysis III. Proceedings of the Third International Symposium on Multivariate Analysis Held at Wright State University, Dayton, Ohio, June 19-24, 1972* New York: Academic Press. 157-169.

James, A.T. (1977). Discussion on "On resolving the controversy in statistical inference" by G.N. Wilkinson. *J. Roy. Statist. Soc. Ser. B* **39**, 157.

Jessor, R., Graves, T.D., Hanson, R.C. and Jessor, S.L. (1968). *Society, Personality and Deviant Behaviour: A Study of a Tri-ethnic Community.* New York: Holt, Rinehart and Winston.

Kale, B.K. (1964). A note on the loss of information due to grouping of observations. *Biometrika* **51**, 495-497.

Kale, B.K. (1966). Approximations to the maximum likelihood estimator using grouped data. *Biometrika* **53**, 282–285.

Kamien, M. (1975a). A survey of drug use in a part-aboriginal community. *Med. J. Aust.* **1**, 261–264.

Kamien, M. (1975b). Aborigines and alcohol – intake, effects and social implications in a rural community in Western New South Wales. *Med. J. Aust.* **1**, 291–298.

Kamien, M. (1978). The measurement of alcohol consumption in Australian aborigines. *Comm. Health Stud.* **2(3)**, 149–151.

Khavari, K.A. and Farber, P.D. (1978). A profile instrument for the quantification and assessment of alcohol consumption: The Khavari alcohol test. *J. Stud. Alc.* **39(9)**, 1525–1539.

Kirsch, A.D., Newcomb, C.H. and Cisin, I.H. (1965). *An Experimental Study of Sensitivity and Survey Techniques in measuring drinking practices.* Washington D.C., Social Research Project, Report No. 1 (George Washington University).

Knupfer, G. and Room, R. (1964). Age, sex and social class as factors in amount of drinking in a metropolitan community. *Social Problems* **12(2)**, 224–240.

Krige, D.G. (1960). On the departure of ore value distributions from the lognormal model in South African gold mines. *J. Sth. African Inst. Min. Metall.* **61**, 231–244.

Krige, D.G. (1961). Developments in the valuation of gold mining properties from borehole results. *Trans. 7th Comm. Min. Metall. Congress, Johannesburg*, 537–561.

Krupinski, J. (1978). Measurement of alcoholic consumption in Victorian surveys. *Comm. Health Stud.* **2(3)**, 140–144.

Krupinski, J., Baikie, A.S., Stoller, A., Graves, J., O'Day, D.M. and Polke, P. (1967). A community health survey of Heyfield, Victoria. *Med. J. Aust.* **June 17**, 1204–1211.

Krupinski, J., Stoller, A., Baikie, A.G. and Graves, J.E. (1970). *A community health survey of the rural town of Heyfield, Victoria, Australia.* Melbourne: Mental Health Authority, Special Publication No 1.

Kulldorff, G. (1961). *Estimation from grouped and partially grouped samples.* New York: John Wiley and Sons.

Larntz, K. (1978). Small-sample comparison of exact levels for chi-squared goodness-of-fit statistics. *J. Amer. Statist. Assoc.* **73**, 253–263.

Ledermann, S. (1946). La mortalité des adults en France. *Population* **1**, 662–668.

Ledermann, S (1948). La surmortalité des hommes en France. *Cahiers Francais d'Information* **118**, 18–23.

Ledermann, S. (1952a). Influence de la consommation de vins et d'alcools sur les cancers la tuberulose pulmonaire et sur d'autres maladies. *Semaine Medicale* **28**, 221–235.

Ledermann, S (1952b). Une mortalité d'origine économique en France: la mortalité d'origine ou d'appoint alcoolique. *Semaine Medicale* **28**, 417-421.

Ledermann, S. (1952c). Une évaluation du nombre des alcooliques en France depuis 1945. *Population* **7**, 227-236.

Ledermann, S (1953). L'alcoolisation excessive et la mortalité des Francais. *Concours Medicale* 1485-1496,1583-1598,1675-1676,1767-1774.

Ledermann, S. (1956). *Alcool, alcoolisme, alcoolisation. Données scientifiques de caractère physiologique, économique et social.* Paris: Presses Universitaires de France. Cahier no. 29, 314p.

Ledermann, S. (1964a). *Alcool, alcoolisme, alcoolisation. Mortalité, morbidité, accidents du travail.* Paris: Presses Universitaires de France. Cahier no. 41. 613p.

Ledermann, S. (1964b). Kann man den Alkoholismus ohne gleichzeitige Änderung des Gesamtverbrauches einer bevölkerung, reduzieren? *27 Internationaler Kongress, Alkohol und Alkoholismus; Deutscher Hauptstelle gegen die suchtgefahren.* 99-104.

Ledermann, S. and Tabah, F. (1951). Nouvelles données sur la mortalité d'origine alcoolique. *Population* **6**, 41-58.

Link, R.F. and Koch, G.S. (1975). Some consequences of applying lognormal theory to pseudolognormal distributions. *Math. Geol.* **7(2)**, 117-128.

Little, R.E., Schultz, F.A. and Mandell, W. (1977). Describing alcohol consumption; a comparison of three methods and a new approach. *J. Stud. Alc.* **38(3)**, 554-562.

Lucas, G.H.W., Kalow, W., McColl, J.D., Griffith, B.A. and Smith, H.W. (1953). Quantitative studies of the relationship between alcohol levels and motor vehicle accidents. *Proc. 2nd. Int. Conf. Alc. Road Traffic.* 139–145.

Lynch, P., Woodward, D., Waters, M., Kirk, I., Maclean, A., Rockcliffe, W. and Reid, P. (1981). Alcohol consumption of Tasmanian high school pupils. *Aust. Paediatr. J.* **17,** 24–28.

Mäkelä, K. (1969). *Alcoholinkulutuksen jakautuma; kulutustutkimuksen ennakkoraportti (The distribution of use of alcohol; a preliminary report of the consumption study).* Helsinki: Social Research Institute for Alcohol Studies (mimeograph).

Mäkelä, K. (1971a). Alkoholinkulutuksen jakautama (The distribution of the use of alcohol). *Alkoholikysymys* **39,** 3–13.

Mäkelä, K (1971b). *Measuring the consumption of alcohol in the 1968–1969 alcohol consumption study*. Helsinki: Social Research Institute of Alcohol Studies, (rep. no. 2)

Mäkelä, K. (1971c). Concentration of alcohol consumption. *Scandinavian Studies in Criminology* **3,** 77–88.

Mäkelä, K. (1978). Level of consumption and social consequences of drinking. In: **Israel, Y., Glaser, F.B., Kalant, H., Popham, R.E., Schmidt, W. and Smart, R.G. (eds.)** *Research Advances in Alcohol and Drug Problems, Vol. 4.* New York: Plenum Press, 303–348.

Mandell, W. (1982). Preventing alcohol-related problems and dependencies through information and education programs. Chapter 37 in: **Pattison, E.M.**

and **Kaufman, E.** (eds.) *Encyclopedic Handbook of Alcoholism.* New York: Gardner Press 468-482.

Maxwell, M.A. (1952). Drinking in the state of Washington. *Quart. J. Stud. Alc.* **13**, 219-239.

McCall, M.G., Cullen, K.J. and Wearne, K.L. (1978). Measurement of alcohol consumption in individuals in the population of Busselton, Western Australia. *Comm. Health Stud.* **2(3)**, 145-148.

McCarroll, J.W. and Haddon, W. Jr. (1962). A controlled study of fatal automobile accidents in New York City. *J. Chron. Dis.* **15**, 811-826.

McDermott, D. and Scheurich, J. (1977). The logarithmic normal distribution in relation to the epidemiology of drug abuse. *Bull. on Narcotics* **29(1)**, 13-19.

Michelini, C. (1972). Convergence patterns of the scoring method in estimating parameters of a log-normal function. *J. Amer. Statist. Soc.* **67**, 319-323.

Midanik, L. (1982). The validity of self-reported alcohol consumption and alcohol problems: a literature review. *Brit. J. Addict.* **77**, 357-382.

Miller, G.H. and Agnew, N. (1974). The Ledermann model of alcohol consumption. Description, implications and assessment. *Quart. J. Stud. Alc.* **35**, 877-898.

Miller, P.M., Ingham, J.G., Plant, M.A. and Miller, T.I. (1977). Alcohol consumption and self disclosure. *Brit. J. Addict.* **72**, 296-300.

Millwood, J.E. and Mackay, A. (1978). Measurement of alcohol consumption in the Australian population. *Comm. Health Stud.* **2(3),** 123-132.

Mohan, D., Sharma, H.K., Sundaram, K.R. and Neki, J.S. (1980). Pattern of alcohol consumption or rural Punjab males. *Ind. J. Med. Res.* **72,** 702-711.

Moler, C. (1976). *Matlab Users' Guide.* Dept. Computer Science, University of New Mexico.

Moser, J. (1980). *Prevention of Alcohol-related Problems. An International Review of Preventative Measures, Policies and Programmes.* Toronto, Canada: Alcohol and Drug Addiction Research Foundation.

Mulford, H.A. (1964). Drinking and deviant drinking, U.S.A., 1963. *Quart. J. Stud. Alc.* **25,** 634-650.

Mulford, H.A. and Miller, D.E. (1960). Drinking in Iowa. II. The extent of drinking and selected sociocultural categories. *Quart. J. Stud. Alc.* **21,** 26-39.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalised linear models. *J. Roy. Statist. Soc. Ser. A* **135,** 370-384.

Nilsson, T. and Svensson, P.G. (1971). *Ungdomens alkoholvanor och alkoholattityder. (The drinking habits of youth and their attitudes toward alcohol).* Statens offentliga utredningar (Government Official Reports) 1971:77, Stockholm.

O'Connell, B., Hudson, R. and Graves, G. (1979). Alcohol and Tobacco Use. In: **Krupinski, J. and Mackenzie, A. (eds.)** *The Health and Social Survey of the North West Region of Melbourne.* Melbourne: Institute of

Mental Health Research and Postgraduate Training, Mental Health Divn., Health Commission of Victoria (special publication no. 7). 96-111, 265-270.

O'Neill, P. (1977). Grogcount. The quantitative measurement of ethanol intake. *Aust. J. Alcohol and Drug Dependence*. **4(4)**, 109-111.

O'Neill, P. (1979). Alconfrontation. In: *1979 Autumn School of Studies on Alcohol and Drugs*. Melbourne: St. Vincent's Hospital. 113-115.

O'Neill, B. and Wells, W.T. (1971). Blood alcohol levels in drivers not involved in accidents and the lognormal distribution. *Quart. J. Stud. Alc.* **32**, 798-803.

Parker, D.A. and Harman, M.S. (1978). The distribution of consumption model of prevention of alcohol problems. A critical assessment. *J. Stud. Alc.* **39**, 377-399.

Pernanen, K (1974). Validity of survey data on alcohol use. In: **Gibson, R.J. *et al* (ed.)** (1974). *Research advances in alcohol and drug problems Vol. 1*. New York: John Wiley and Sons. 355-374.

Phillips, E.L.R, Little, R.E., Hillman, R.S., Labbe, R.F. and Campbell, C. (1984). A field test of the sweat patch. *Alcoholism: Clin. Exp. Res.* **8**, 233-237.

Phillips, M. (1982). Sweat-patch test for alcohol consumption: rapid assay with an electrochemical detector. *Alcoholism: Clin. Exp. Res.* **6**, 532-534.

Phillips, M. (1984). Sweat-patch testing detects inaccurate self-reports of alcohol consumption. *Alcoholism: Clin. Exp. Res.* **8**, 51-53.

Plant, M.A. and Miller, T-I. (1977). Disguised and undisguised questionnaires compared: Two alternative approaches to drinking behaviour surveys. *Social Psychiatry*. **12,** 21-24.

Popham, R.E. and Schmidt, W. (1981). Words and deeds: the validity of self-report data on alcohol consumption. *J. Stud. Alc.* **42,** 355-368.

Popham, R.E., Schmidt, W. and De Lint, J. (1971). Epidemiologic research bearing on legislative attempts to control alcohol consumption and alcohol problems. In: *Symposium on Law and Drinking Behaviour*. Chapel Hill, N.C.: Center for Alcohol Studies, University of North Carolina, 4-16.

Popham, R.E., Schmidt, W. and de Lint, J. (1976). The effects of legal restraint on drinking. Chapter 13 in: **Kissin, B. and Begleiter, H. (eds.)** (1976). *Social aspects of alcoholism*. New York: Plenum Press

Quensel, C.-E. (1945). Studies of the logarithmic normal curve. *Skand. Aktuartidsk*. **28,** 141-153.

Quinn, K., Costello, B. and Rigby, K. (1975). *Drug use among Adelaide Secondary School Students: An interim report*. South Australia: S.A. Foundation for Alcohol and Drug Dependence.

Rankin, J.G. (1971). The size and nature of the misuse of alcohol and drugs in Australia. In: **Kiloh, L.G. and Bell, D.S. (eds.)** *29th International Congress on Alcoholism and Drug Dependence, Sydney, Australia, February 1970*. Australia: Butterworths, 11-20.

Rankin, J.G. (1974). *The politics of alcohol and drug use*. 7th Leonard Ball Oration. Melbourne: Victorian Foundation for Alcoholism and Drug Dependence.

Rankin, J.G. and Wilkinson, P. (1971). Alcohol and tobacco consumption. In: **Krupinski, J. and Stoller, A. (eds.)** (1971). *The Health of a Metropolis.* Australia: Heineman.

Rao, C.R. (1952). *Advanced Statistical Methods in Biometrical Research.* New York: John Wiley and Sons.

Rao, C.R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankya* **18**, 139-148.

Rasumovsky, N.K. (1940). Distribution of metal values in ore deposits. *C.R. Acad. Sci. U.R.S.S.* **28**, 814.

Reynolds, I., Gallagher, H. and Bryden, D. (1977). Heavy drinking and regular drug taking amongst a sample of Sydney adults. *Aust. J. Alcohol and Drug Dependence.* **4(4)**, 125-127.

Reynolds, I., Harris, J., Gallagher, H. and Bryden, D. (1976). Drinking and drug taking patterns of 8516 adults in Sydney. *Med. J. Aust.* **2**, 782-785.

Robinson, D. (1982). Alcoholism: perspectives on prevention strategies. Chapter 36 in: **Pattison, E.M. and Kaufman, E. (eds.)** *Encyclopedic Handbook of Alcoholism.* New York: Gardner Press, 458-467.

Room, R. (1971). Drinking laws and drinking behaviour: some past experience. In: *Symposium on Law and Drinking Behaviour.* Chapel Hill, N.C.: Center for Alcohol Studies, University of North Carolina. 29-55.

Ross, G.J.S. (1980). *MLP - Maximum Likelihood Program.* Harpenden, U.K.:

Rothamsted Experimental Station.

Ryan, G.A. and Salter, W.E. (1977). *Blood alcohol levels and drinking behaviour of road crash victims*. Melbourne: Department of Social and Preventive Medicine, Monash University.

Saker, B.M., Tofler, O.B., Burvill, M.J. and Reilly, K.A. (1967). Alcohol consumption and gout. *Med. J. Aust.* **1**, 1213-1216.

Sargent, M. (1979). *Drinking and Alcohol in Australia*. Melbourne: Longman Chesire.

Schact, P., Brown, J., Brown, R. and Schonfeid, C. (1976). *The Redcliffe Health Survey*. Brisbane: Queensland Department of Health.

Schmidt, W. (1973). Analysis of alcohol consumption data; the use of consumption data for research purposes. In: *The epidemiology of dependencies: report of a conference*. Copenhagen: WHO Regional Office for Europe. 57-66.

Schmidt, W. and De Lint, J. (1970). Estimating the prevalence of alcoholism from alcohol consumption and mortality data. *Quart. J. Stud. Alc.* **31**, 957-964.

Schmidt, W. and Popham, R.E. (1978). The single distribution theory of alcohol consumption: A rejoinder to the critique of Parker and Harman. *J. Stud. Alc.* **39(3)**, 400-419.

Selge, B. (1975). *The South East Regional Health Survey*. S.A.: S.A. Department of Public Health.

Sichell, H.S. (1947). An experimental and theoretical investigation of bias error in mine sampling with special reference to narrow gold reefs. *Trans. Inst. Min. Metall. (London)* **56**, 403-473.

Singh, G. (1979). Comment on "The single distribution theory of alcohol consumption". *J. Stud. Alc.* **40(5)**, 522-524.

Skog, O.-J. (1971). *Alkoholkonsumets fordeling i befolkningen (the distribution of alcohol in the population).* Oslo: National Institute for Alcohol Research.

Skog, O.-J. (1973). Less alcohol - fewer alcoholics? *Drinking and Drug Practices Surveyor* **7**, 7-14.

Skog, O.-J. (1974). *Bidrag til en teori om alkoholkonsumets fordeling II (A contribution to the theory of the distribution of alcohol consumption).* Oslo: National Institute for Alcohol Research.

Skog, O.-J. (1977a). On the distribution of alcohol consumption. In: *The Ledermann Curve. Report of a Symposium held in London on 6 and 7 January 1977*. London: The Alcohol Education Centre, 25-43.

Skog, O-J. (1977b). *Does the same distributional model for alcohol consumption apply to both male and female populations?* Oslo: National Institute for Alcohol Research, mimeograph series no. 10.

Skog, O.-J. (1979a). A note on the distribution of alcohol consumption: a reply and rejoinder. *The Drinking and Drug Practices Surveyor* **14**, 3-6.

Skog, O.-J. (1979b). The distribution of alcohol consumption: a further reply to Guttorp and Song. *Drinking and Drug Practices Surveyor* **15**, 23-

24.

Skog, O.-J. (1979c). *Modeller for Drikkeatferd (Models for Drinking Behaviour).* Oslo: National Institute for Alcohol Research, mimeograph series no. 32.

Skog, O.-J. (1980a). Total alcohol consumption and rates of excessive use: a rejoinder to Duffy and Cohen. *Brit. J. Addict.* **75,** 133-145.

Skog, O.-J. (1980b). Is alcohol consumption lognormally distributed? *Brit. J. Addict.* **75,** 169-173.

Skog, O.-J. (1980c). Social interaction and the distribution of alcohol consumption. *J. Drug Issues.* **10(1),** 71-92.

Skog, O.-J. (1981). Alcoholism and social policy: are we on the right lines? *Brit. J. Addict.* **76,** 315-321.

Skog, O.-J. (1982). *The distribution of alcohol consumption. Part I: A critical discussion of the Ledermann model.* Oslo: National Institute for Alcohol Research. mimeographed series no. 64.

Skog, O.-J. (1983). *The distribution of alcohol consumption. Part II: A review of the first wave of empirical studies.* Oslo: National Institute for Alcohol Research. mimeographed series no. 67.

Smart, R.G. (1977). Social policy and the prevention of drug abuse: Perspectives on the unimodal approach. Chapter 10 in: **Glatt, M.M. (ed.)** *Drug Dependence. Current Problems and Issues.* Lancaster, U.K.: MTP Press Ltd.

Smart, R.G. (1978). The distribution of illicit drug use: correlations between extent of use, heavy use and problems. *Bull. on Narcotics* **30(1)**, 33-41.

Smart, R.G. and Liban, C.B. (1982). Alcohol consumption as estimated by the informant method: a household survey and sales data. *J. Stud. Alc.* **43(9)**, 1020-1027.

Smart, R.G. and Schmidt, W. (1970). Blood alcohol levels in drivers not involved in accidents. *Quart. J. Stud. Alc.* **31**, 968-971.

Smart, R.G. and Whitehead, P.C. (1972). The consumption patterns of illicit drugs and their implications for prevention of abuse. *Bull. on Narcotics* **24(1)**, 39-47.

Smart, R.G. and Whitehead, P.C. (1973). The prevention of drug abuse by lowering *per capita* consumption: Distributions of consumption in samples of Canadian adults and British university students. *Bull. on Narcotics* **25(4)**, 49-55.

Smart, R.G., Whitehead, P.C. and Laforest, L. (1971). The prevention of drug abuse by young people: an argument based on the distribution of drug abuse. *Bull. on Narcotics* **23(2)**, 11-15.

Smith, N.M.H. (1976a). *Ledermann Procedure*. Australia: Senate Standing Committee on Social Welfare (Reference: Continuing oversight of the drug use problem). Official Hansard transcript of the evidence, Vol 1, 348-364.

Smith, N.M.H. (1976b). Research note on the Ledermann formula and its recent applications. *Drinking and Drug Practices Surveyor* **12**, 15-22.

Stacey, B.G. and Elvy G.A. (1981). Is alcohol consumption log-normally distributed among fourteen- to seventeen-year-olds? *Psych. Rep.* **48**, 995-1005.

Storm, T. and Cutler, R.E. (1981). Observations of drinking in natural settings. Vancouver beer parlors and cocktail lounges. *J. Stud. Alc.* **42(11),** 972-997.

Straus, R. and Bacon, S.D. (1953). *Drinking in College*. New Haven: Yale University Press.

Strömbom, M. (1975). *The Bradbury Health and Welfare Survey*. Parramatta, Western Metropolitian Health Region, Health Commission of NSW.

Sudman, S. (1980). Reducing response error in surveys. *Statistician* **29(4),** 237-273.

Sulkunen, P. (1977). Behind the curves: on the dynamics of rising consumption level. In: *The Ledermann Curve. Report of a symposium held in London on 6 and 7 January 1977*. London: The Alcohol Education Centre, 44-57.

Swamy, P.S. (1960). Estimating the mean and variance of a normal distribution from singly and doubly truncated samples of grouped observations. *Cal. Statist. Assoc. Bull.* **9,** 145-156.

Swamy, P.S. (1962). On the amount of information supplied by censored samples of grouped observations in the estimation of statistical parameters. *Biometrika* **49,** 245-249.

Swamy, P.S. (1963). On the amount of information supplied by truncated samples of grouped observations in the estimation of the parameters of normal populations. *Biometrika* **50,** 207-213.

Tallis, G.M. and Young, S.S.Y. (1962). Maximum likelihood estimation of parameters of the normal, log-normal, truncated normal and bivariate normal distributions from grouped data. *Aust. J. Statist.* **4**, 49-54.

Taylor, C. (1979). A method for describing variability in alcohol consumption levels. *Brit. J. Addict.* **74**, 57-66.

Tofler, O.B., Saker, B.M., Rollo, K.A., Burvill, M.J. and Stenhouse, N. (1969). Electrocardiogram of the social drinker in Perth, Western Australia. *Brit. Heart J.* **31**, 306-313.

Tofler, O.B. and Woodings, T.L. (1981). A 13-year follow-up of social drinkers. *Med. J. Aust.* **2**, 479-481.

Tuck, M. (1980). *Alcohol and Social Policy: are we on the right lines?* London: H.M.S.O.

Turner, T.J. and McClure, L. (1975). *Alcohol and drug use by Queensland school children.* Queensland: Department of Education, Research Branch.

Vamosi, M. (1960). Determination of the amount of alcohol in the blood of motorists. *Traffic Safety* **4**, 8-11.

Volicer, B.J. and Volicer, L. (1982). Randomised response technique for estimating alcohol use and noncompliance in hypertensives. *J. Stud. Alc.* **43(7)**, 739-750.

von Mises, R. (1918). Uber die 'Ganzzahligheit' der Atomgewichte und Verwandte Fragen. *Physikalishe Zeitschrift* **7**, 153-159.

Whittacker, E.T. and Robinson, G. (1932). *The calculus of observations. A* *treatise on numerical mathematics.* London: Blackie and Son Ltd.

Wilkinson, G.N. (1977). On resolving the controversy in statistical inference (with Discussion). *J. Roy. Statist. Soc. Ser. B* **39,** 119–171.

Wilkinson, P., Santamaria, J.N., Rankin, J.G. and Martin, D. (1969). Epidemiology of alcoholism: social data and drinking patterns of a sample of Australian alcoholics. *Med. J. Aust.* **1,** 1020–1025.

Wilson, P. (1981). Improving the methodology of drinking surveys. *Statistician* **30,** 159–167.

World Health Organisation. (1974). *WHO Expert Committee on Drug Dependence. Twentieth Report.* Technical Report Series no. 551. Geneva, W.H.O.