

CPH
FAST



**A STATISTICAL STUDY OF
THE DISTRIBUTION OF ALCOHOL CONSUMPTION
AND CONSEQUENT INFERENCEAL 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

Adelaide

June 1985

CONTENTS

Summary	vi
Signed statement	viii
Acknowledgements	ix
Chapter 1: Introduction.	1

PART I

Chapter 2: The distribution of alcohol consumption – an historical overview.	4
2.1 Ledermann's original proposals, 1956	4
2.2 The period 1968 – 1975	10
2.3 The report of Bruun <i>et al</i> , 1975	17
2.4 The period since 1975	21
2.5 Discussion	27
Chapter 3: The Ledermann model of alcohol consumption.	34
3.1 Introduction	34
3.2 The Ledermann procedure and the Ledermann model	35
3.2.1 Overview	35
3.2.2 Description	35
3.2.3 Summary	38
3.3 The Ledermann model as a reparameterisation of the two parameter lognormal distribution	39
3.4 Characterisation	41
3.5 Ledermann's data	43
3.6 The value of θ	45
3.7 The value of the maximal consumption	48
3.8 An example	50
3.9 Discussion	54

Chapter 4: Other models of alcohol consumption.	59
4.1 The two parameter lognormal distribution	59
4.1.1 Definition	59
4.1.2 Characteristics	59
4.1.3 The proportion of heavy consumers	60
4.2 The three parameter lognormal distribution	64
4.3 Truncated and censored lognormal distributions	65
4.4 Estimation of lognormal distributions from grouped data	66
4.4.1 Introduction	66
4.4.2 Maximum likelihood estimation of lognormal distributions from grouped data – a brief review	66
4.4.3 Details necessary for maximum likelihood fitting of lognormal distributions using iterated weighted regression	68
4.5 The gamma distribution	73
4.5.1 Definition	73
4.5.2 Characteristics	73
4.5.3 The proportion of heavy drinkers	74
4.6 A model relating age subpopulations	76
Chapter 5: Australian data on the distribution of alcohol consumption.	78
5.1 Methods of measuring individual alcohol consumption	78
5.1.1 Introduction	78
5.1.2 Present consumption	79
5.1.3 Past consumption	80
5.2 Units of measurement of alcohol consumption	85
5.3 The validity of survey data on alcohol consumption	89
5.4 The data	93

Chapter 6: Results.	121
6.1 Scope of analyses	121
6.2 Abstainers	123
6.3 Consumers - sample statistics	125
6.3.1 Sample sizes and groupings	125
6.3.2 Mean consumption and standard deviation	127
6.3.3 Skewness	128
6.4 Fitted distributions	129
6.4.1 Introduction	129
6.4.2 Fits to aggregate adult groups	131
6.4.3 Fits to age subgroupings	133
6.4.4 The relation between the parameters of the two parameter lognormal fits	136
6.4.5 Comparison of censored and truncated lognormal fits	136
6.4.6 Comparison of censored and uncensored lognormal fits	138
6.5 Mean consumption and proportion of heavy consumers	140
6.6 Discussion	145
Appendix: Details of fitted distributions.	148

PART II

Chapter 7: Inference on linear functions of class probabilities.	181
7.1 Introduction	181
7.2 The choice of a specification for the distribution of alcohol consumption	185
7.3 Linear functionals relevant to alcohol studies	191
7.4 Linear algebra for estimation from grouped data - preliminaries	194
7.4.1 Basic definitions and notation	194
7.4.2 Asymptotic assumption	198
7.4.3 Maximum likelihood estimation as iterated weighted regression	197
7.5 Sample and contrast spaces	200
7.5.1 Sample space	200
7.5.2 Contrast space	201
7.5.3 Inner product metrics and identity transforms	202
7.5.4 The score-functional subspace of contrast space	204
7.5.5 Orthogonal decompositions of sample and contrast space	204

7.6	Decomposition theorem	207
7.7	Partition of contrasts in parametric estimation	212
7.7.1	Introduction	212
7.7.2	Partitions of contrasts	213
7.7.3	Partitions of χ^2	216
7.7.4	Partitions of variance	217
7.7.5	Example	217
7.7.6	Discussion	220
7.8	Modifications of the two parameter lognormal distribution: a comparison of adding a third parameter and censoring the lower tail	222
7.8.1	Introduction	222
7.8.2	Relationship of the three parameter and censored two parameter lognormal distributions to the two parameter lognormal distribution	223
7.8.3	Approximations to the three parameter and censored two parameter lognormal distributions	226
7.8.4	Comparison of the distributions via the covariates	229
7.8.5	The removal of spurious information, part 1	229
7.8.6	Further decomposition of linear functionals	231
7.8.7	The removal of spurious information, part 2	233
7.8.8	Discussion and summary	239
7.9	Fitting a distribution subject to a constraint on a linear function of the fitted probabilities	242
7.9.1	Introduction	242
7.9.2	Fitting the model	243
7.9.3	Example	245
References	247

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.