Monographs
on Statistics and
Applied Probability 46

Analysis of Quantal Response Data

B.J.T. Morgan

CHAPMAN & HALL

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

D.R. Cox, D.V. Hinkley, D.B. Rubin and B.W. Silverman

1 Stochastic Population Models in Ecology and Epidemiology *M.S. Bartlett* (1960)

2 Queues *D.R. Cox and W.L. Smith* (1961)

3 Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)

4 The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)

5 Population Genetics *W.J. Ewens* (1969)

6 Probability, Statistics and Time *M.S. Bartlett* (1975)

7 Statistical Inference *S.D. Silvey* (1975)

8 The Analysis of Contingency Tables *B.S. Everitt* (1977)

9 Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)

10 Stochastic Abundance Models *S. Engen* (1978)

11 Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)

12 Point Processes *D.R. Cox and V. Isham* (1980)

13 Identification of Outliers *D.M. Hawkins* (1980)

14 Optimal Design *S.D. Silvey* (1980)

15 Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)

16 Classification *A.D. Gordon* (1981)

17 Distribution-free Statistical Methods *J.S. Maritz* (1981)

18 Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)

19 Applications of Queueing Theory *G.F. Newell* (1982)

20 Risk Theory, 3rd edition *R.E. Beard, T. Pentikainen and E. Pesonen* (1984)

21 Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)

22 An Introduction to Latent Variable Models *B.S. Everitt* (1984)

23 Bandit Problems *D.A. Berry and B. Fristedt* (1985)

24 Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)

25 The Statistical Analysis of Compositional Data *J. Aitchison* (1986)

26 Density Estimation for Statistical and Data Analysis *B.W. Silverman* (1986)

27 Regression Analysis with Applications *G.B. Wetherill* (1986)

28 Sequential Methods in Statistics, 3rd edition *G.B. Wetherill* (1986)

29 Tensor Methods in Statistics *P. McCullagh* (1987)

30 Transformation and Weighting in Regression *R.J. Carroll and D. Ruppert* (1988)

31 Asymptotic Techniques for Use in Statistics *O.E. Barndoff-Nielson and D.R. Cox* (1989)

32 Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)

33 Analysis of Infectious Disease Data *N.G. Becker* (1989)

34 Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)

35 Empirical Bayes Method, 2nd edition *J.S. Maritz and T. Lwin* (1989)

36 Symmetric Multivariate and Related Distributions *K.-T. Fang, S. Kotz and K. Ng* (1989)

37 Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)

38 Cyclic Designs *J.A. John* (1987)

39 Analog Estimation Methods in Econometrics *C.F. Manski* (1988)

40 Subset Selection in Regression *A.J. Miller* (1990)

41 Analysis of Repeated Measures *M. Crowder and D.J. Hand* (1990)

42 Statistical Reasoning with Imprecise Probabilities *P. Walley* (1990)

43 Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)

44 Inspection Errors for Attributes in Quality Control *N.L. Johnson, S. Kotz and X. Wu* (1991)

45 The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46 Analysis of Quantal Response Data *B.J.T. Morgan* (1992)

(Full details concerning this series are available from the Publishers.)

Analysis of Quantal Response Data

B. J. T. MORGAN
Professor of Applied Statistics
Institute of Mathematics and Statistics
University of Kent
Canterbury

CHAPMAN & HALL
London · Glasgow · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN

Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Blackie Academic & Professional, Wester Cleddens Road, Bishopbriggs, Glasgow G64 2NZ, UK

Chapman & Hall, 29 West 35th Street, New York NY10001, USA

Chapman & Hall Japan, Thomson Publishing Japan, Hirakawacho Nemoto Building, 6F, 1-7-11 Hirakawa-cho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, Thomas Nelson Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1992

Typeset in 10/12 Times by Thomson Press (India) Ltd, New Delhi
Printed in Great Britain by St. Edmundsbury Press,
Bury St. Edmunds, Suffolk

ISBN 0 412 31750 8

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication data available

Morgan, Byron J. T., 1946–
Analysis of quantal response data / B. J. T. Morgan. – 1st ed.
p. cm. – (Monographs on statistics and applied probability; 46)
Includes bibliographical references and indexes.
ISBN 0–412–31750–8
1. Biometry. 2. Probits. 3. Medical statistics. 4. Quantum statistics. I. Title. II. Series.
QH323.5.M67 1992 92–19056
574'.01'5195–dc20 CIP

'All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy.'

Paracelsus (1493–1541)

Contents

Preface

This book has grown out of a lecture course on biometry given to M.Sc. students in statistics at the University of Kent. The standard reference for the course was the book *Probit Analysis* by Professor D. J. Finney. It is now 20 years since the appearence of the 3rd edition of *Probit Analysis* and there have been many developments in statistics of relevance for the analysis of quantal response data during this time, in design, sequential methods, non-parametric procedures, over-dispersion, robust methods, Bayesian approaches, extended models, influence and diagnostics, synergy and many other areas. The single most important development is probably the introduction of generalized linear models, allied to specialist computer packages for fitting these models. Most computer packages now provide a menu of relevant procedures for quantal assay data. Additionally the whole computing scene has changed dramatically, with the move towards powerful personal computers and workstations. The aim of this book has been to describe the new developments for the analysis of quantal response data, and to emphasize the links between the various different areas.

Several extra-mural courses have been given on the text material. The first of these was at the Royal Melbourne Institute of Technology, given jointly with Professor R. G. Jarrett. The last was at Duphar, in Weesp in The Netherlands, and in between two residential courses were given at the University of Kent. The interest shown by the course participants was one motivation for writing this book.

Quantal response data are quite often used to illustrate statistical techniques, and readers of the book will find that they will encounter many different areas of statistics. The book may be read by people with a range of different backgrounds. It is designed to be read as a coherent text or as a source of reference. Numerate scientists should

be able to follow many of the arguments. However, for a full understanding a mathematics background is necessary. Much of the material should be accessible to third year mathematics and statistics undergraduates in British universities who have had foundation and second-level courses in statistics in their first two years of study. The book should be ideal for study at the postgraduate level by students of statistics and biometry. There are 267 Exercises to help with the use of the book as a course text.

The first four Appendices help to make the book complete, and the fifth summarizes useful computing facilities. For illustration, a number of GLIM macros appear in the text, and a small number of examples are given in BASIC and MINITAB. However, prior knowledge of these packages/languages is not a prerequisite for understanding the material of the book.

Well over 50 data sets are presented. Several of these now have **classic** status, in that they are, sometimes uncritically, regularly used to illustrate new procedures. Some of the examples have arisen from consulting experience with the Division of Animal Health at CSIRO, Melbourne, with Pfizer Central Research, Sandwich, Kent, and with Shell Research, Sittingbourne, Kent.

I am grateful to many individuals for their help and comments while this book has been written. At the Biometry Division of Pfizer, Kent, a range of problems were raised and discussed by P. Colman, R. Hews, T. Lewis, H. Ross-Parker and R. White. I have been particularly fortunate in supervising two postgraduate students working in relevant areas. The material of Chapters 5 and 6 owes a clear debt to the Ph.D. thesis of Simon Pack, who worked on a CASE award with Wellcome Research Laboratories, Beckenham, Kent, while Chapter 4 has likewise benefited from the M.Sc. dissertation of Paul Goedhart.

Deborah Ashby read the entire book as referee, and both Martin Ridout and David Smith read particular chapters. To these three I am most grateful for many helpful corrections and comments. Prominent amongst the others whom I should thank are: Beverley Balkau, John Fenlon, Janneke Hoestra, Hans Jansen and Richard Jarrett. Michael Bremner provided computing advice and help with the troff system. Encouragement was provided by Professors B. M. Bennett and A. A. Rayner, and the late David Williams, and useful advice by Sir David Cox. Parts of the book were typed by Lilian Bond and Arija Crux but the lion's share of the labour was

carried out by Mavis Swain, who surpassed even her legendary typing skills with great humour and patience.

Finally I thank my wife, Janet, and children, Chloë and Leo, for their tolerant acceptance of my regular weekend absences over the four-year period when the book was written.

Byron J. T. Morgan
Canterbury

Glossary and notation

Except where noted below, a standard notation is used throughout the book. A small number of notational clashes have been adopted between different chapters if that improved comprehension or corresponded to standard usage.

corpus luteum: glandular tissue in the ovary, which forms after rupture of the follicle at ovulation. It secretes progesterone.

dominant lethal test: experiment in which experimental units are male animals, and each male is mated to one or more females.

fecundability: probability of conception per menstrual cycle.

implant: used here to denote egg implanted in womb following fertilization.

isolate: a pure culture of an organism.

micromelia: abnormally small size of the arms or legs.

minimum inhibitory concentration: lowest concentration (of an anti-infective agent) at which a particular organism's growth is inhibited.

phocomelia: congenital absence of the upper arm and/or upper leg (e.g. as side-effect of thalidomide).

E[ ]: expectation.
V( ): variance.
Pr( ): probability.
L: likelihood.
l: log-likelihood.
D: deviance.
X^2: Pearson goodness-of-fit statistic.
$N(\mu, \sigma^2)$: normal distribution, mean μ, variance σ^2.
$\Phi(x)$: standard normal c.d.f.
$\phi(x)$: standard normal p.d.f.

$\Phi(x, y)$: bivariate standard normal c.d.f.

ED_{100p} (LD_{100p}, EC_{100p}) $= \theta_p$ (for $p \neq 0.5$); $ED_{50} = \theta$ (but see also beta-distributions, below).

$\hat{\theta}_R$: Reed–Muench estimator of θ; $\hat{\theta}_D$: Dragstedt–Behrens estimator of θ.

E_M, E_B, E_{DB}, estimators of θ from the up-and-down experiment.

k: number of doses, cases or treatments.

m: number of signal presentations (Chapter 3)/number of sampling times (Chapter 5)/number of litters in a group (Chapter 6).

$\{d_i, 1 \leqslant i \leqslant k\}$: doses.

[More generally $\{x_i\}$, or $\{z_i\}$, when doses are not involved, or a transformation is used.]

$\Delta_i = (d_{i+1} - d_i)$.

n_i individuals are treated at dose d_i and r_i respond.

In the time interval $(0, t_j)$, r_{ij} respond to dose d_i; $n_{ij} = r_{ij} - r_{i,j-1}$ (Chapter 5).

Of n_{ij} insects exposed to a_i units of A and b_j units of B, r_{ij} die (Chapter 3).

$P_i = P(d_i) =$ probability of response to dose d_i/probability that x_j respond out of n_j in the mixture model of equation (6.12).

$\mathbf{P}_i$: transition matrix for movement between states between times t_{j-1} and t_j (Chapter 5).

$p_i = r_i/n_i$ (Chapter 2).

$P(t|d) =$ probability of response to dose d by time t.

$p_{ij} =$ probability of response to dose d_i in time-interval (t_{j-1}, t_j).

$s_i = r_i - n_i P_i$.

$$\tilde{s}_i = \frac{r_i - n_i \hat{P}_i}{\sqrt{n_i \hat{P}_i (1 - \hat{P}_i)}}$$

t: used to denote time/iterate number, as in $\zeta^{(t)}$.

Bin (n, p): binomial distribution, index n and probability p.

$B(\alpha, \beta)$: beta function: $\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$.

$I_{k,\zeta}(x)$: incomplete gamma integral.

μ: mean of a random variable, especially (Chapter 6) for a beta random variable.

$\hat{\mu}$: Spearman–Kärber estimate of θ (estimate of mean of tolerance distribution).

$\hat{\mu}_\alpha$: $\alpha\%$ trimmed Spearman–Kärber estimate.

μ_j: jth non-central moment of tolerance distribution.

$\mu_{(j)}$: jth central moment of tolerance distribution.
h: number of hits in multi-hit model.
$h(\mu)$, $h(\mu; \lambda)$: link functions.
$h_v(q)$: quantits.
$R(\alpha)$: Mills ratio.
$F(x)$: cumulative distribution function (c.d.f.).
$\hat{F}(x)$: empirical c.d.f.
$\tilde{F}(x)$: estimate of c.d.f. by linear interpolation from ABERS estimate.
tr(**A**): trace of matrix **A**.
I(v, ζ), **J**(v, ζ); Fisher information/expected information matrices.
$\psi(u)$: kernel function (Chapter 7).
$\phi_{T,F}(x)$: influence curve.
$SC_{T,m}(x)$: Tukey's sensitivity curve.
$IC_{T,F,D}(d, y)$: influence curve based on dose mesh D.
$IC_{T,F}(d, y)$: influence curve.
(α, β): location-scale pair of parameters/standard parameterization for beta distribution.
(μ, θ): alternative parameterization for beta distribution (but note also use of θ for ED_{50}).
$\rho = \theta/(1 + \theta)$, and also more generally as correlation between littermates/also ρ is used to denote relative potency.
m_i: number of litters in ith treatment group.
n_{ij}: size of jth litter in ith treatment group.
x_{ij}: number responding out of n_{ij} (Chapter 6).
p_i: probability of response for all litters in ith treatment group (Chapter 6).
[NB: subscript i is sometimes dropped for simplification.]
$\mathbf{X} = \{x_{ij}\}$, the design matrix (except Chapter 6).

Index of data sets

CHAPTER 1

Data, preliminary analyses and mechanistic models

1.1 Introduction

The names Hiroshima, Chernobyl and Thalidomide are synonymous with twentieth century tragedies. Many years after the exploding of atomic bombs over Hiroshima and Nagasaki, the effects of exposure to nuclear radiation on the survivors are quantifiable in terms of increased incidence of leukaemia, as described in the paper by Armitage and Doll (1962). An example involving chromosome aberration in survivors of Hiroshima is considered in Chapter 6. The drug **Thalidomide** had been prescribed as a safe hypnotic drug, but the winter of 1961 saw the horrifying reports of its use resulting in babies born with deformities of **phocomelia**, or **micromelia**. As Beedie and Davies (1981) ominously wrote, 'It had not been tested in animals for teratogenicity, but thousands of babies born to mothers who had taken the drug during pregnancy provided the missing data.'

Common to these two illustrations is the exposure of human beings to substances that are either unnatural, or provided at unnaturally high levels. The response of individuals, adults or embyros, is binary: they are either affected by the time they are inspected, or they are not. In more general terms, discrete responses may take a variety of forms, such as reduction of pain, alleviation of breathing problems, improvement in acne, remission from leukaemia, and so on. Data which quantify the effect of exposure of individuals to substances such as new drugs, or to radiation, are often described as discrete, or quantal. Responses need not just be binary, and later we shall see examples of quantal data which may result in three or more possible outcomes.

This book is concerned with the analysis of quantal response data, sometimes called dose-response data, or quantal assay data. Such

data may arise in a wide variety of different areas as we shall see, and may be collected from a properly designed scientific experiment, or result from observational studies. Thus, for example, girls of different ages may be classified by whether or not they have started menstruation; patches of woollen fabrics may be assessed for the degree of 'prickle' they elicit in human subjects; viruses ingested by insects may or may not kill them; widely used food additives may be tested for their undesirable side-effects. An example of this was cited in an article in *The Independent* newspaper of 23 September 1987: 'Several preservatives may cause asthmatic reactions in susceptible people. And one, *methyl paraben* (E218), is the main volatile component in the vaginal secretions of beagles – it may cause socially embarrassing behaviour in dogs. E218 is used in beer and coffee and many other foods.'

The main emphasis in the examples which illustrate the methodology of the book will be on the evaluation and testing of substances, mainly drugs, for use in humans. Frequently the effect investigated is whether or not there is a positive outcome from using the drugs resulting in efficacy studies; however, it is also of vital importance to consider the possible harmful side-effects of otherwise potentially beneficial treatments. Thus, for example, patients suffering from the spine-fusing disease, Ankylosing Spondylitis, may be treated by radiation therapy, but leukaemia may result as an undesired side-effect (Court Brown and Doll, 1957). The radiation used in mammography has been estimated as likely to cause just one excess cancer per 10^6 million women screened (Whitehouse, 1985; see also Breslow and Day, 1980, p. 62). The Babylonian Code of Hammurabi, of 2200 BC, ordained that if a patient died, the treating physician should lose his hands, and this is regarded as the first example, indeed a somewhat extreme one, of the need for regulation of procedures for treating human beings. In modern times, pharmacopoeias have been devised throughout the world, presenting standards for drug purity. The first statute to control drug quality in America was passed in 1848, while as recently as 1968 the Medicines Act of Great Britain produced new safeguards for the development, production and use of new drugs.

Because a number of the examples in this book are drawn from toxicology, it is worthwhile outlining important aspects of toxicology before we start, and this is done in the next section. An excellent

introduction to the statistical aspects of the full range of drug development and testing is given by Salsburg (1990).

1.2 Aspects of toxicology

The activity of chemical substances can sometimes be gauged from their physico-chemical properties, and the Quantitative Structure–Activity Relationship (QSAR) procedures described by Bergman and Gittins (1985) are designed to search for new active substances using physical structure and electrochemical property correlates with established substances of known performance.

New chemicals may also be tested *in vitro*. Thus for example the Ames test for mutagenicity positively identified 157 out of a series of 175 known carcinogens (McCann *et al.*, 1975). Ultimately, however, tests *in vivo* are necessary. The revolutionary oral anti-fungal drug fluconazole was not found to be especially effective *in vitro*: the 'modest *in vitro* profile understates the excellent *in vivo* activity of fluconazole demonstrated in animal models of fungal infections and in clinical trials' (Marriott and Richardson, 1987). See also Exercise 1.26. In using non-human animals as models for humans the basic assumption is always that the model is appropriate. With the possible exception of arsenic, all known chemical carcinogens in humans are carcinogenic in some, but not all species of animals used in laboratories, so the model has to be chosen with care (Klaassen, 1986). For further discussion on the extrapolation from animals to humans, see Mantel and Bryan (1961), Cornfield (1977) and Park and Snee (1983).

Carageenan, which is a seaweed extract, is used in products such as ice cream and biscuits, yet it causes changes resembling ulcerative colitis in the bowels of guinea pigs, rabbits and mice. Inevitably effects such as these are the result of doses given at far higher levels than those commonly encountered in foods, and to relatively small groups of animals. This is a standard toxicological procedure, and is necessary in order to reduce cost and unnecessary suffering in experimental animals. The difficult problem is then to extrapolate from a known dangerous dose in animals to a virtually safe one for human consumption, and we discuss this fundamental problem in sections 1.6 and 4.6.

Different toxins may be administered in different ways, for example

Table 1.1

	Weight (g)	*Dosage (mg/kg)*	*Dose (mg/animal)*	*Surface area (cm^2)*	*Dosage (mg/cm^2)*
Mouse	20	100	2	46	0.043
Rat	200	100	20	325	0.061
Guinea pig	400	100	40	565	0.071
Rabbit	1500	100	150	1270	0.118
Cat	2000	100	200	1380	0.145
Monkey	4000	100	400	2980	0.134
Dog	12000	100	1200	5770	0.207
Human	70000	100	7000	18000	0.388

through ingestion, by contact with the skin, or by intravenous injection, and their effect can be radically affected by the size of the animal tested. Thus it is quite usual for dosages to be given in mg/kg of body weight, for example, or mg/cm^2 of body area. Table 1.1, taken from Klaassen (1986), shows how a constant dosage measured in mg/kg translates into different overall doses per animal, for a variety of species, and different dosages in terms of mg/cm^2. It is difficult to appreciate what a dosage measured in mg/kg actually becomes when scaled up to life-size, and Table 1.2, also taken from Klaassen (1986), provides the required interpretation, together with a crude toxicity rating to describe the different lethal doses. The distinction between dose and dosage that is drawn here will be maintained throughout the book.

Before new drugs can be tested in the standard progression of clinical trials on human subjects, they may be screened on a variety

Table 1.2

Toxicity rating or class	*Probable lethal oral dose for humans*	
	Dosage	*For average adults*
1. Practically nontoxic	> 15 g/kg	More than 1 quart
2. Slightly toxic	5–15 g/kg	Between pint and quart
3. Moderately toxic	0.5–5 g/kg	Between ounce and pint
4. Very toxic	50–500 mg/kg	Between teaspoonful and ounce
5. Extremely toxic	5–50 mg/kg	Between 7 drops and teaspoonful
6. Supertoxic	< 5 mg/kg	A taste (fewer than 7 drops)

of animals, with tests designed for a corresponding range of different effects. These include acute and chronic toxicity, with experiments in the latter case possibly running for a number of years, and usually performed on rats. Rabbits are the preferred animal for tests for eye and skin irritation, while the guinea pig is usually used for tests for skin sensitization, when this seems appropriate. Tests for possible teratological effects usually involve rats and rabbits, and substances are administered to males and/or females, before mating, and, for females, during gestation, and during lactation. Observations include the pregnancy rate and the viability of progeny, and study may continue for several generations. Mutation effects can be sought through a number of *in vivo* and *in vitro* procedures. The dominant lethal test, which we encounter again in Chapter 6, involves giving a male animal (usually a rodent) a single dose of the compound prior to mating with one or two females. The females are then killed before term, and numbers of live embryos and *corpora lutea* recorded for analysis. The extrapolation from animals to humans takes us through what has been referred to as the 'species barrier'.

We see that substances may be administered in a variety of ways, and by single or repeated doses. Substances which are toxic by one route of application may not prove toxic by another: the skin may prove to be an effective barrier to poisons; the liver may detoxify a substance given orally, which may be far more toxic if inhaled, for example. While a compound itself may not be toxic, a metabolite of it might be. Clearly tests must try to reflect the intended use of substances. If they are likely to find their way into water, they need to be assessed for possible effects on fish, crustacea and so forth. Aquatic experiments may differ from those on mammals in that exposure to the toxic agent may be continuous.

Many of the features described in this section will be encountered in the examples which now follow.

1.3 Examples

We shall now present a number of examples to illustrate the wide range of problems to be considered, and to provide instances of the different types of experiment described in the last section. In all cases response is quantal, and in most cases there is a single covariate, such as mean age group or dose level, which is deemed likely to affect the response. In some cases there are several covariates, which

may, singly or in conjunction, influence the response. We shall see also the kinds of questions that arise and require answers in an appropriate statistical setting. The remainder of this chapter also serves as an introduction to the rest of the book.

Example 1.1

An experiment to assay an anti-pneumococcus serum (dose measured in cc).

Irwin (1937) analysed the data of Table 1.3. Groups of mice were given a serum inoculation, at various doses, prior to being infected with pneumococci.

We see that as the dose of serum is increased, the proportion of mice protected increases. The relationship between dose and resulting proportion is frequently simplified by transformation in each case. Here we have logarithms of doses, and the commonly used transformation, $\text{logit}(p) = \log_e\{p/(1-p)\}$ of proportions. Of interest here is the serum level to set for routine anti-pneumococcus inoculation.

Table 1.3 *Effect of anti-pneumococcus serum on mice*

$10.158 + \log_2$ (serum dose)	No. of mice protected	No. of mice in experiment	Proportion protected (p)	Logit (p)
1	0	40	0.000	–
2	2	40	0.050	−2.944
3	14	40	0.350	−0.619
4	19	40	0.475	−0.100
5	30	40	0.750	1.099

Example 1.2 Age of menarche in 3918 Warsaw girls

This example differs from most of the others presented in this chapter in that the data arise from an **observational** study rather than an experimental one. However, we can see the qualitative similarity between the data of Tables 1.3 and 1.4, and we shall see later how they may be analysed by the same methods. Nevertheless there remains an important distinction between the two different types of study, and we shall at times find it necessary to emphasize this distinction.

Table 1.4 *Age of menarche in Polish girls*

Mean age of group (years)	*No. having menstruated*	*No. of girls*	*Proportion having menstruated (p)*	*Logit (p)*
9.21	0	376	0.000	–
10.21	0	200	0.000	–
10.58	0	93	0.000	–
10.83	2	120	0.017	−4.076
11.08	2	90	0.022	−3.784
11.33	5	88	0.057	−2.809
11.58	10	105	0.095	−2.251
11.83	17	111	0.153	−1.710
12.08	16	100	0.160	−1.658
12.33	29	93	0.312	−0.792
12.58	39	100	0.390	−0.447
12.83	51	108	0.472	−0.111
13.08	47	99	0.475	−0.101
13.33	67	106	0.632	0.541
13.58	81	105	0.771	1.216
13.83	88	117	0.752	1.110
14.08	79	98	0.806	1.425
14.33	90	97	0.928	2.554
14.58	113	120	0.942	2.781
14.83	95	102	0.931	2.608
15.08	117	122	0.959	3.153
15.33	107	111	0.964	3.283
15.58	92	94	0.979	3.829
15.83	112	114	0.982	4.025
17.58	1049	1049	1.000	–

These data were presented by Milicer and Szczotka (1966) and record, for a sample of 3918 Warsaw girls taken in 1963, whether or not they had reached menarche (onset of menstruation). This is probably the best known of a number of studies of age of menarche. Other studies include those by Burrell *et al.* (1961) and Milicer (1968). Data resulting from the second of these papers are presented in Exercise 2.23. Interestingly, differences are detectable between individuals of different race and of different socio-economic status. From a purely statistical point of view, in the **experimental** context, data sets as large as these are less frequently encountered than much smaller sets, such as that of Table 1.3, and may allow discrimination between competing simple probability models which usually are indistinguishable.

Example 1.3 The effect of insecticide on flour-beetles

Hewlett (1974) observed the effect of insecticide sprayed onto flour-beetles at four different concentrations. The data given in Table 1.5 differ from those of Table 1.3 in that insects are used, application is topical, by spraying, different sexes are distinguished and also the observations are made at a number of times, rather than just one. The data of the last two rows present the responses for the entire length of the experiment, or endpoint mortalities as they are called, and so are qualitatively similar to the data of Table 1.3.

When presented with such data we might look for sex differences, both in terms of overall response and speed of response. When summarizing overall responses rates, or when comparing these between sexes, we might question the extent to which precision and

Table 1.5 *Numbers of male (M) and female (F) flour-beetles (Tribolium castaneum) dying in successive time intervals following spraying with insecticide (Pyrethrins B) in Risella 17 oil. The beetles were fed during the experiment in an attempt to eliminate natural mortality. Data from Hewlett (1974)*

Time interval (days)	*Concentration (mg/cm² deposit)*							
	0.20		0.32		0.50		0.80	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*
0–1	3	0	7	1	5	0	4	2
1–2	11	2	10	5	8	4	10	7
2–3	10	4	11	11	11	6	8	15
3–4	7	8	16	10	15	6	14	9
4–5	4	9	3	5	4	3	8	3
5–6	3	3	2	1	2	1	2	4
6–7	2	0	1	0	1	1	1	1
7–8	1	0	0	1	1	4	0	1
8–9	0	0	0	0	0	0	0	0
9–10	0	0	0	0	0	0	1	1
10–11	0	0	0	0	0	0	0	0
11–12	1	0	0	0	0	1	0	0
12–13	1	0	0	0	0	1	0	0
No. survivors	101	126	19	47	7	17	2	4
No. treated	144	152	69	81	54	44	50	47

power have been increased by collecting data over time. We consider this point in detail in Chapter 5.

The beetles involved here are *Tribolium castaneum*, the rust-red flour-beetle. They are small insects, 3–4 mm long, infesting flour, and eating this or broken grain (Hewlett, P. S., personal communication). The fact that the insects were sprayed means that different beetles receive different doses, for a given concentration. The analysis in Chapter 5 ignores this feature, but it is discussed in section 3.9.

Example 1.4 Recovery of insects

An important feature of aerosol fly sprays is whether they knock flies down, and not necessarily whether the flies are actually killed in the process – sometimes flies recover from 'knock-down', as the data of Table 1.6 show. How might we compare the results of the two experiments? We discuss a mechanistic model for such data in Chapter 5.

Example 1.5 Experiments to investigate the effect of arboviruses on chicken eggs

Jarrett *et al.* (1981) analysed experiments carried out to investigate the effects of arboviruses injected into chicken embryos. The aim was to quantify the potency of arboviruses, with a view ultimately to assessing how these might affect lamb foetuses. Two examples of the resulting data are given in Table 1.7.

In this example there are three possible responses, and, as was implicit also in the last two examples, we are interested in comparisons between sets of data. Data of this kind frequently result from making observations over time, as in the last two examples, but the time information is suppressed in this case. Thus in Table 1.7 eggs were classified 18 days after injection of the virus; non-specific deaths in the first few days were excluded, each group of eggs having been originally of size 20. An illustration of time-dependent data for this kind of experiment is given in Table 1.8. Eggs were candled, i.e. held up to the light, each day to see whether the embryo was dead or alive.

In many investigations responses may be due to different causes. Presented with pairs of different woollen fabrics, with only one of each pair being 'prickly', subjects who cannot discriminate between

Table 1.6 *For two experiments, A and B, the data below give the numbers of houseflies (Musca domestica) airborne at several times after the initial dose of spray was administered: a fixed amount of spray was released into a wind tunnel in which the flies were allowed to fly freely. Data from Pack (1986a)*

	Experiment					
	A			*B*		
	Concentration (μg/l)			*Concentration (μg/l)*		
Time (minutes)	0.3	1.0	2.0	0.3	1.0	2.0
1	18	12	9	19	19	10
5	15	0	0	10	0	0
10	12	0	0	12	0	0
20	15	2	0	13	0	0
60	18	4	0	18	13	0
180	18	16	17	20	22	10
group size	18	16	22	20	22	20

Table 1.7 *The effect of two arboviruses on chicken embryos*

Virus	*Inoculum titre (PFU/egg)*	*No. of eggs*	*Dead*	*Alive*	
				Deformed	*Not deformed*
Facey's	3	17	3	1	13
Paddock	18	19	4	1	14
	30	19	8	2	9
	90	20	17	1	2
Tinaroo	3	19	1	0	18
	20	19	2	0	17
	2400	15	4	9	2
	88000	19	9	10	0
Control		18	1	0	17

the fabrics by touch may correctly identify the prickly item by chance. In other cases the correct response can result from a clear perception of prickle on the part of the subjects. Death may result from a cause other than the application of a poison. Even onset of menstruation may, in some cases, be incorrectly ascribed to bleeding due to

Table 1.8 *Time course of an experiment to investigate the effect of an arbovirus on chicken embryos. The data give the cumulative number dead out of 20, except for log dose 0.65, when an egg was dropped on day 8*

	Day													
Log dose	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.65	0	0	1	1	3	3	3	3	4	4	4	4	4	4
2.50	2	2	2	2	2	2	2	3	3	3	3	3	4	4
4.32	2	2	2	2	2	2	2	4	4	6	7	9	9	11
6.23	0	1	1	1	2	2	3	6	7	10	11	12	12	14
Control	0	0	1	1	1	1	1	1	1	1	1	1	1	1

pathological causes. In Example 1.3, beetles were fed in order to minimize natural mortality. In cases where natural response is possible, it is advisable for control groups to be employed, as in Table 1.8. Further illustrations are given in Example 1.7. Ways of dealing with natural response as in a control group are considered in section 3.2.

Example 1.6 Hypersensitivity reactions to a drug

The data in Table 1.9 are taken from a much larger study into the possible side-effects of a drug. Differing experimental protocols at different sites resulted in experiments of appreciably variable lengths

Table 1.9 *Hypersensitivity reactions to a drug, administered at four sites, A, B, C or D*

Site				*Time on drug (days)*	*Presence of a reaction (1 = reaction)*	*Sex (2 = female)*	*Dose (mg)*
A	*B*	*C*	*D*				
1	0	0	0	11	1	2	250
1	0	0	0	22	0	2	250
1	0	0	0	20	0	1	250
1	0	0	0	7	1	1	100
0	0	0	1	78	0	2	250
0	0	0	1	27	0	2	50
0	1	0	0	399	0	1	150
0	0	1	0	55	0	1	125

being run before the studies were terminated. Here, as in the last two examples, times are recorded in addition to whether a response took place. Of primary importance to the pharmaceutical company involved was whether there was evidence of hypersensitivity reactions being related to the dose level used.

Example 1.7 Foetal death in a control population

New drugs need to be tested carefully for any possible effects on pregnant animals. The data in Table 1.10 are taken from Haseman and Soares (1976) and just describe control groups from dominant lethal assays, mentioned in section 1.2. In this experiment a drug's ability to cause damage to reproductive genetic material, sufficient to kill the fertilized egg or developing embryo, is tested by dosing a male mouse and mating it to one or more females. A significant

Table 1.10a *Sample No. 1 of Haseman and Soares (1976)*

	Observed frequency distribution of foetal death in mice													
Litter size	*Number of dead foetuses* 0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2													
2	2													
3	3													
4	5	1	1											
5	2	2												
6	2	2												
7	2	2	2	1										
8	6	1		1	1									
9	2	3	1											
10	2	4	2		2									
11	19	11	3	3										
12	33	24	11	5	4	4							1	
13	39	27	12	6	5	2			1					
14	34	30	14	6	6			1						
15	38	22	18	4	2	1								
16	13	16	14	4	3	1								
17	8	4	3	3	2	1		1						
18		4	2	1										
19	2	1												
20														1

Table 1.10b *Sample No. 3 of Haseman and Soares (1976)*

	Observed frequency distribution of foetal death in mice									
Litter size	*Number of dead foetuses* 0	1	2	3	4	5	6	7	8	9
1	7									
2	7									
3	6									
4	5	2	1							
5	8	2	1		1	1				
6	8									
7	4	4	2	1						
8	7	7	1							
9	8	9	7	1	1					
10	22	17	2		1			1	1	
11	30	18	9	1	2		1		1	
12	54	27	12	2	1		2			
13	46	30	8	4	1	1		1		
14	43	21	13	3	1					1
15	22	22	5	2	1					
16	6	6	3		1	1				
17										
18	3		2	1						

increase in foetal deaths is then indicative of a mutagenic effect. We need to consider how we might describe such data sets in a relatively simple manner, and how we might make comparisons with similar data sets corresponding to treated animals. This is the topic of Chapter 6, where the basic assumption of a binomial distribution for responses is relaxed to accommodate **extrabinomial variation**, which usually arises when different litters of animals are involved.

Example 1.8 Signal detection experiments

A common experiment in psychology involves presenting subjects with stimuli which may either just be noise (N), or may involve a signal superimposed upon noise (SN). The subjects indicate whether or not they thought the signal was present, sometimes qualifying their responses with a measure of confidence. The performance of subjects may be monitored under a variety of adverse environmental conditions, and it is then of interest to measure the extent to which

Table 1.11 *Data resulting from a signal detection experiment on three different subjects. Each subject responds 'Yes' if the stimulus was thought to be present, 'No' if it was not, etc.*

		Responses		
Subject	*Stimulus*	*No*	*Not sure*	*Yes*
1	N	30	10	15
	SN	9	7	35
2	N	25	17	5
	SN	2	18	30
3	N	18	7	3
	SN	2	10	16

performance may change as conditions change. Such experiments may model behaviour such as the vigilance of radar screen monitors in submarines. The data of Table 1.11, taken from Grey and Morgan (1972), provide an illustration.

Again we want to summarize the data, and make comparisons between subjects. Relevant analysis is provided in section 3.5. An alternative form of signal detection experiment arises when subjects are informed that a stimulus is present on just one of m occasions, for $m > 2$, and the subjects have to select the occasion they think corresponds to the signal presentation. This is called an m-alternative forced-choice experiment, and will also be discussed again in Chapter 3.

Example 1.9 The Australian bovine tuberculosis eradication campaign

In work aimed at the eradication of bovine tuberculosis in Australia, suitably treated bovine tissue is placed on culture plates and examined for the growth of colonies of *Mycobacterium bovis*. Material for culture is decontaminated prior to inoculation onto culture media and the data in Table 1.12 describe colony counts when two different decontaminants (HPC and oxalic acid) are applied, in varying concentrations. While there are obvious similarities between this experiment and, say, that of Example 1.1, there is no universal upper limit to a colony count, and the data of Table 1.12

Table 1.12 *Colony count data, taken from Trajstman (1989). The value marked by a * was omitted from all analyses*

[HPC] % weight/ volume	*Decontaminant: HPC No.* M. Bovis *colonies at stationarity*										*Sample mean*	*Sample variance*
0.75	2	4	8	9	10	1	0	5	14	7	6.0	19.6
0.375	11	12	13	12	11	13	17	16	21	2	12.8	24.4
0.1875	16	6	20	23	23	39	18	23	33	21	22.2	80.6
0.09375	33	46	42	18	35	20	19	29	41	36	31.9	102.3
0.075	30	30	27	53	51	39	31	36	38	22	35.7	100.0
0.0075	53	62	38	54	54	38	46	58	54	57	51.4	66.5
0.00075	3*	42	45	49	32	39	40	34	45	51	41.9	40.6

[Oxalic acid] % weight/ volume	*Decontaminant: oxalic acid No.* M. Bovis *colonies at stationarity*										*Sample mean*	*Sample variance*
5	14	15	6	13	4	1	9	6	12	13	9.3	23.1
0.5	27	33	31	30	26	41	33	40	31	20	31.2	39.1
0.05	33	26	32	24	30	52	28	28	26	22	30.1	70.8
0.005	36	54	31	37	50	73	44	50	37	–	45.8	164.4

Control experiment where no decontaminant is used No. M. Bovis *colonies at stationarity*										*Sample mean*	*Sample variance*
52	80	55	50	58	50	43	50	53	54	51.8	110.8
44	51	34	37	46	56	64	51	67	40		

will require different probability models, to be described in sections 3.3 and 6.5.2. See also Exercise 1.22.

Example 1.10 Serological data

The results of a serological survey carried out in Zaire into the extent of malarial infection in individuals aged greater than 6 months are given in Table 1.13. In this example the percentage sero-positive is bounded above by a factor reflecting the overall incidence of malaria. We consider modelling these data in Chapter 3.

Table 1.13 *Data from Bongono (Zaire) showing the proportions of individuals in different age groups with antibodies present, as assessed by a particular serological test. Data from Marsden (1987)*

Mean age group (years)	*No. of individuals examined*	*No. sero-positive*	*Percentage sero-positive*
1.0	60	2	3.3
2.0	63	3	4.8
3.0	53	3	5.7
4.0	48	3	6.3
5.0	31	1	3.3
7.3	182	18	9.9
11.9	140	14	10.0
17.1	138	20	14.5
22.0	84	20	23.8
27.5	77	19	24.7
32.0	58	19	32.8
36.8	75	24	32.0
41.6	30	7	23.3
49.7	62	25	40.3
60.8	74	44	59.5

Example 1.11 Mixtures of drugs

The data of Table 1.14 result from an experiment designed to investigate how two insecticides (A and B) may act in combination. Of interest here is whether insecticides interact to produce enhanced performance (synergy), or a reduction in performance (antagonism). An analysis of these data is provided by Giltinan *et al.* (1988) and we discuss their findings in section 3.7.

Table 1.14 *The results of a study to investigate the contact insecticidal activity of mixtures of two insecticides, A and B. The target insect was the tobacco budworm, Heliothis virescens. Treatment was administered by means of direct application of one microlitre for each dosage to the body of each insect. Mortality was measured 96 hours after treatment (Data from Giltinan et al., 1988)*

Mixture	*Amount of A(ppm)*	*Amount of B(ppm)*	*Number of dead insects*	*Number of insects tested*
B	0	30.00	26	30
B	0	15.00	19	30
B	0	7.50	7	30
B	0	3.75	5	30
A25:B75	6.50	19.50	23	30
A25:B75	3.25	9.75	11	30
A25:B75	1.625	4.875	3	30
A25:B75	0.812	2.438	0	30
A50:B50	13.00	13.00	15	30
A50:B50	6.50	6.50	5	30
A50:B50	3.25	3.25	4	29
A50:B50	1.625	1.625	0	29
A75:B25	19.50	6.50	20	30
A75:B25	9.75	3.25	13	30
A75:B25	4.875	1.625	6	29
A75:B25	2.438	0.813	0	30
A	30.00	0	23	30
A	15.00	0	21	30
A	7.50	0	13	30
A	3.75	0	5	30

1.4 Preliminary graphical representations

An obvious first approach to the kind of data illustrated so far is to plot proportions affected against dose, or log dose, or time, or whatever appears appropriate. This is done in Figures 1.1–1.3 for the data in Tables 1.3–1.5, respectively. The value of doing this is illustrated in Figure 1.3, for example: we can appreciate that males appear to be more susceptible than females and, furthermore, that when they respond they appear to do so more quickly than females. We shall quantify these differences by using mixture models, from the area of survival analysis, in Chapter 5.

One may well consider fitting a straight line to points such as those of Figure 1.1. However, it is preferable to transform the

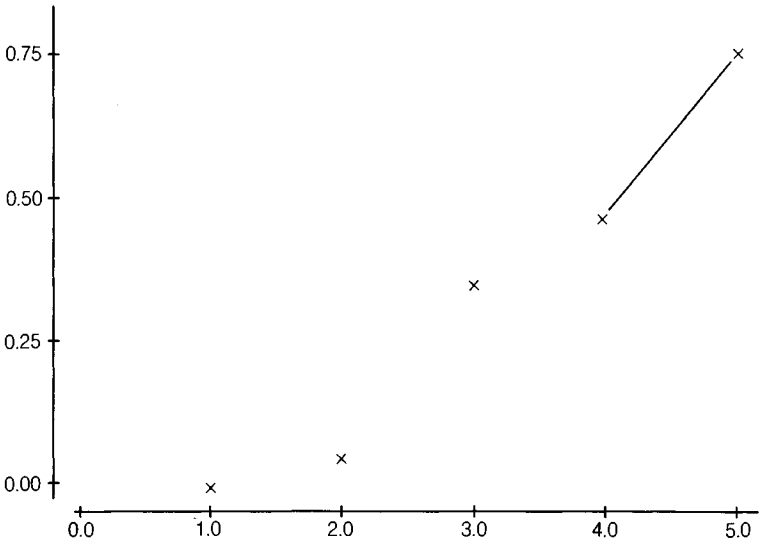

Figure 1.1 *A plot of the proportions protected versus* $\log_2$ *(dose) for the data of Table 1.3. The reason for connecting the two proportions shown is given in section 1.6.*

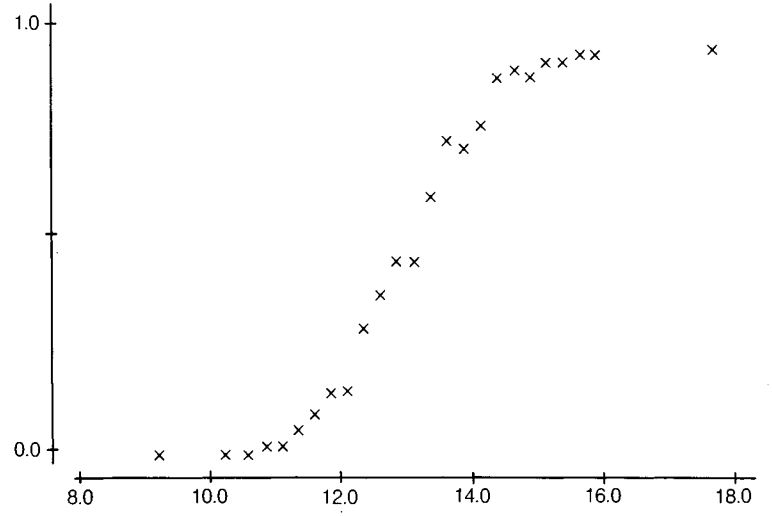

Figure 1.2 *A plot of the proportions of Table 1.4 versus mean age of groups.*

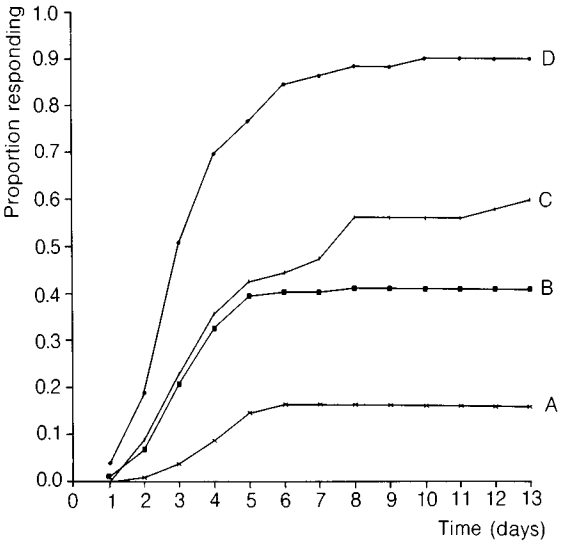

Figure 1.3a *A plot of the proportions of female beetles responding versus time, from Table 1.5, reproduced from Pack (1986a).*

Key	
0.2 mg/cm^2	A
0.32 mg/cm^2	B
0.50 mg/cm^2	C
0.80 mg/cm^2	D

proportions first. In many cases the plot corresponding to that of Figure 1.1 has a more sigmoid appearance, as is true of the points of Figure 1.2. If we plot the logits of the proportions versus age for the data of Table 1.4, we obtain the more linear plot of Figure 1.4. Finite logits do not exist for proportions of 0 or 1. Corresponding doses are indicated by arrows on the graph. Special graph paper may be used if the plotting is to be done by hand.

For Figure 1.4 the least squares linear regression line is,

$$\text{logit}\{p(x)\} = -20.8 + 1.58x \tag{1.1}$$

where $p(x)$ denotes the proportion that have reached menarche by age x. The product–moment correlation between logit $\{p(x)\}$ and x has value 0.992, and so one might feel that the data are well described by equation (1.1). However, the proportions of Figure 1.2 result from

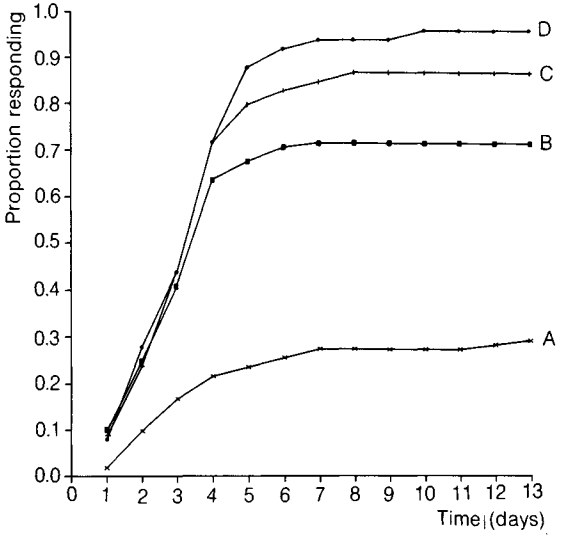

Figure 1.3b *A plot of the proportions of male beetles responding versus time, from Table 1.5, reproduced from Pack (1986a).*

Key	
0.2 mg/cm^2	A
0.32 mg/cm^2	B
0.50 mg/cm^2	C
0.80 mg/cm^2	D

binomial distributions, and will have unequal variances. A weighted regression, weighting inversely with respect to the variance of logit (p) (Exercise 1.2) gives the regression line:

$$\text{logit}\{p(x)\} = -20.0 + 1.54x \qquad (1.2)$$

revealing little difference from equation (1.1) – but see also Exercise 1.3. The fitted line of equation (1.1) is called a minimum chi-square line, while that of equation (1.2) results from the technique of minimum logit chi-square. These methods are compared in section 2.6, after a full discussion of the maximum-likelihood estimation procedure for standard quantal response data.

Before fitting lines to data we can also consider whether a preliminary transformation of the explanatory variate, x, might improve the fit. A logarithmic transformation is often used routinely

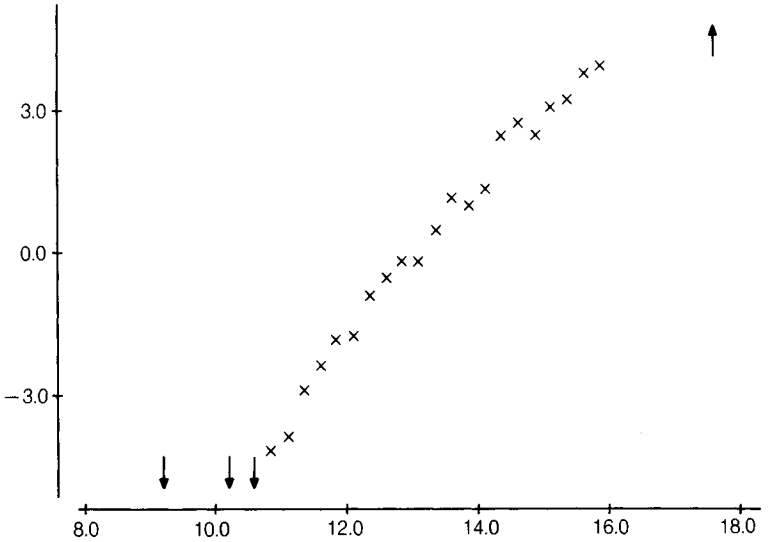

Figure 1.4 *A plot of the logits of the proportions of Table 1.4 versus age, excluding points for which the logit is not finite.*

for this purpose, though in some cases it has impaired the fit, rather than improved it; we shall discuss this further in detail in Chapter 4. In many exploratory investigations of new substances, their potency may be uncertain before the experiment. In such a case a wide range of dose levels is therefore sensible, and a natural device is to space the doses equally on a log scale, and then also for convenience to present results, plots and analyses in terms of that scale.

If we set $x = 0$ in equation (1.2) we see that birth equates to onset of mentruation in a very small, but non-zero, fraction of newborn female children. This should not worry us unduly, since it involves extrapolation well outside the age range over which data were collected. Over the age range of the collected data the model may provide a succinct description of the data. However, we can see that if the model had been formulated in terms of log(age), this problem would not have arisen. In fact, as we shall see in Chapter 4, the logarithmic transformation improves the fit of the model for these data. Extrapolation is the subject of section 1.6. The use of logarithms can also appear naturally from various mechanistic models which we shall now describe.

1.5 Mechanistic models

When we consider the data of Table 1.4, on age of onset of menstruation, it is natural to suppose that age of menarche, in a homogeneous population of girls, has some distribution, with cumulative distribution function $F(\alpha + \beta x)$, where α, β are a location and scale pair of parameters. In such a case, the probability, P, of, for example, 79 girls out of 98 with mean age of 14.08 years having reached menarche can be approximated by the binomial form

$$P \approx \binom{98}{79} F(\alpha + 14.08\beta)^{79} \{1 - F(\alpha + 14.08\beta)\}^{19}$$

The exact form for P takes into account the interval nature of the data – see Exercise 1.13. An obvious contender for the form of $F(\ )$ is the normal distribution. An example of this is cited by Biometrika tables and refers to the detonation of explosives at varying distances from cardboard discs. In each experiment the proportion of discs perforated is noted (Exercise 1.14).

This same model is also used for standard quantal assay data, as in Table 1.3, and it may be justified by the classical, or threshold model for quantal response data. In this model it is assumed that each individual in the relevant population has a dose tolerance, or threshold, T say, to a particular substance. If the dose administered, d, is greater than T then the individual responds. Otherwise it does not. If the tolerances are distributed throughout the population with distribution $F(\alpha + \beta x)$, say, as in the above illustration, then the probability of individual response to dose d is simply:

$$\Pr(T \leqslant d) = F(\alpha + \beta d)$$

In practice this model may also be used when d is a dosage, rather than a dose (see Exercise 1.15 for further discussion). In this book we shall refer to tolerance/threshold models/distributions. A small fraction of individuals may have high tolerances, giving rise to a positively skewed tolerance distribution. A logarithmic dose transformation might then be advantageous if the model to be fitted assumes a symmetric tolerance distribution.

Historically the favoured form for $F(\ )$ has been normal, resulting in what is called **probit** analysis. The greater simplicity of the cumulative distribution function of the similar logistic distribution has resulted in emphasis now being placed on use of the logistic

distribution and the resulting **logit** analysis, a particular example of logistic regression discussed in section 2.8. However, a computational advantage of the probit model arises if a dose d is observed with error, and we return to this errors-in-variables situation in Chapter 3.

Probit or logit models may be adopted from the pragmatic viewpoint of simply requiring an appropriate description of the data, and this has already been done of course in the fitted model of equation (1.2), without any reference to threshold models. There are cases where the threshold model is not appropriate; in cancer formation, for example, tumours may result from a change to a single cell initially and similarly death may follow from infection by a single virus particle. However, threshold models usually provide a useful way of thinking about the data. We shall consider an extension of the threshold model in Chapter 5, when we model times to response. We may note here that there are also other areas where ideas of thresholds have been found to be useful for analysing data; for example, Anderson and Aitkin (1985), and Exercise 1.24.

In psychology, a similar model provides the justification for signal detection theory which is used to describe data such as those of Example 1.8 (Exercises 1.12 and 2.8). This theory has been given a general setting by McCullagh (1980), as a way of analysing contingency table data with ordered categories; here too the terminology of probit and logit (and other) models is used. The resulting models are discussed in section 3.5, for the analysis of multiple response data, as in Table 1.7.

In general terms we may describe models as **mechanistic** or **descriptive** (in the latter case, Ripley, 1987, prefers the term **convenient**). The latter type of model does not rely on a specification of a mechanism, and simply aims to summarize the data, and provide a framework for inference. Thus we may regard a simple linear regression model as descriptive. The end-product of a mechanistic modelling exercise may be a descriptive model, whose parameters play no role other than fitting the model to the data, and we shall encounter several examples of this.

Puri and Senturia (1972) proposed an elaborate mechanistic model for the way in which insects might attempt to shed insecticide, through a random sequence of losses of random amounts. This model was then used to fit data by supposing that the individual insect hazard rate at any time was a function of the amount of insecticide remaining by that time. We consider this model in detail in

Chapters 4 and 5. One-hit and multi-hit models are also described in Chapters 4 and 5. Originally devised as models of carcinogenicity, these models have quite recently been employed for describing quite general quantal response data – see Rai and Van Ryzin (1981). The basic premise is that individuals exposed to a substance can be likened to a target bombarded with arrows, at an intensity determined by the dose level adopted. In the **one-hit** model it is supposed that a single arrow on target is sufficient to elicit a response. The **multi-hit** model is more stringent in requiring several hits. The **multi-stage** model supposes that various stages have to be completed, either in series or in parallel before a response is obtained. For a comprehensive review, see Kalbfleisch *et al.* (1983). If a toxic response results from at least h hits from arrows arriving in a Poisson process at rate ζd for some fixed (say unit) time, where d corresponds to the dose level, then (see Exercise 1.17) the probability of response is given by:

$$P(d; h, \zeta) = \int_0^d \frac{\zeta^h x^{h-1} e^{-\zeta x} dx}{\Gamma(h)} \quad \text{for} \quad 0 < d < \infty$$

and we can see therefore that this model is equivalent to a threshold model with a gamma-distributed tolerance distribution. This is in fact a limiting case of an extended threshold model presented by Prentice (1976b), based on the $\log F$ distribution, with cumulative distribution function,

$$F(z) = \int_{-\infty}^{z} \frac{e^{wm_1}(1+e^w)^{-(m_1+m_2)}}{B(m_1, m_2)} dw, \quad -\infty < z < \infty$$

The logit model is 'extended' by this model, as it forms the special case of $m_1 = m_2 = 1$. There is full discussion of this and other extended models in Chapter 4.

The Puri–Senturia model can also result in a descriptive model with a positively skewed tolerance distribution (Chapter 5), and this finding can help to explain why a preliminary logarithmic dose transformation followed by analysis based on a model which incorporates a symmetric tolerance distribution may often be a useful procedure (for more discussion, see Chapter 4).

At the other end of the dose-response spectrum from a supposed 'target' is the administration of the substance under test. A fundamental defect of many mechanistic models is that they take

the administered dose as the effective dose, when in fact the dose level may be inaccurately measured, much sprayed dose may miss the insect, and so forth. Van Ryzin and Rai (1987) have produced a mechanistic model of this process, based on a compartmental-model assumption (Exercise 2.33). A simpler assumption is to suppose that the effective dose has a Poisson distribution, with mean value the supposed administered dose. This is the approach adopted by Williams (1965) in a mechanistic model for microbial infections based on the stochastic linear birth and death process. This model, considered also by Morgan and Watts (1980), is different from others considered in this book in that the dose refers to a suspension of bacteria, which by their development can effectively change the dose during the course of the infection (Exercise 1.18). The work of section 3.9 considers how errors in dose measurement may be described and built into the model. Ridout and Fenlon (1991) consider microbial control of insect pests. In this case the pathogen is delivered through food, which for a given concentration of pathogen may even result in some insects receiving no pathogen.

Mechanistic models have been found to be particularly useful in justifying various forms of low dose extrapolation procedures, and we describe how in the next section.

1.6 Interpolation and extrapolation

A partial answer to the question: 'What are models for?' has already been provided in the last section: models can simplify, summarize and provide a basis for inference – as in making comparisons; for example, the topic of section 3.7. Mechanistic models may provide much more, since the parameters of the model might correspond to definite aspects of the supposed mechanism. By simplifying the data, models may be thought of as smoothing out variation, and so fitted models may be used for interpolation. Another major use of models in general is in prediction. This is abundantly clear in areas such as time-series analysis, but it is also the objective of low dose extrapolation, mentioned in section 1.2.

As is discussed in detail in section 2.7, we are frequently interested in estimating dose levels, dosage levels or concentrations which correspond to an average given percentage, $100p$, of individuals responding. Traditionally such levels are denoted by expressions such as ED_{100p} or LD_{100p}. Here E denotes 'effective', and L denotes

'lethal'. In some studies it may be an EC_{100p} level which is estimated, corresponding to a particular **concentration**. An example is provided by Ridout and Fenlon (1991) who also discuss how to convert concentrations to doses. By analogy, a single summary of the age-of-menarche data displayed in Figure 1.2 would be the age by which we would expect 50% of girls to have reached menarche, the median of the threshold distribution.

The values in Table 1.15 (taken from Klaassen, 1986) show how an ED_{50} may be useful in providing a simple indication of the relative toxicity of a range of different chemicals (response here is death in treated animals). A review article is given by Zbinden and Flury-Roversi (1981). There is more discussion of ED_{100p} estimation, with particular reference to interval estimation, in section 2.7. While such values are readily estimated from models, it may be argued that a non-parametric procedure could be more relevant. Thus in Figure 1.1, for example, we could simply use the indicated linear interpolation if we wanted to estimate that dose resulting in a 50% response rate. This simple procedure is the simplest illustration of a trimmed Spearman–Kärber estimate, and we discuss this and other non-parametric methods in Chapter 7.

As an alternative to using the ED_{50} for evaluating the toxicity of substances, the British Toxicology Society (1984) proposed a fixed dose procedure. The approach here is to use four dosages, namely 5, 50, 500 and 2000 mg/kg and test 10 animals at that dosage which

Table 1.15

Agent	ED_{50} *(mg/kg)*
Ethyl alcohol	10 000
Sodium chloride	4 000
Ferrous sulphate	1 500
Morphine sulphate	900
Phenobarbital sodium	150
Picrotoxin	5
Strychnine sulphate	2
Nicotine	1
d-Tubocurarine	0.5
Hemicholinium-3	0.2
Tetrodotoxin	0.10
Dioxin (TCDD)	0.001
Botulinum toxin	0.00001

is judged *a priori* to produce a toxic but not lethal effect. Other dosages may or may not then be used, depending on the observed response. This procedure has been evaluated by van den Heuvel *et al.* (1990) and Whitehead and Curnow (1991). It has been found to be more conservative in general than simply using the ED_{50}, but it requires fewer animals to be tested on average.

While Finney (1971, p. 210) emphasizes 'the evils of extrapolation', extrapolation is frequently required in the context of quantal response data, for reasons explained in section 1.2. A review is given by Armitage (1982). This is clearly an area where different models can give quite different predictions. While a model based on the logistic distribution, say, may provide a satisfactory fit to data, as judged perhaps by a chi-square goodness-of-fit test, significant improvements may result from fitting more complex models such as that of Prentice, described above. Improvements of fit may typically be anticipated in the tail areas, and so be particularly important for extrapolation. In the evaluation of virtually safe doses of possible carcinogens, response rates as low as 10^{-6} or 10^{-8} may be involved (Rai and Van Ryzin, 1981). This is an area to which we shall return at various points throughout the book. In practice a variety of procedures is used in such extreme cases of extrapolation (Exercise 1.19).

1.7 Discussion

The material of this chapter provides an introduction to the work of later chapters. The logit transformation was employed in section 1.4, but many others may also have been used, e.g. the probit transformation, with

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx$$

which has slightly lighter tails than the logistic form, the sine curve, with

$$F(z) = \{1 + \sin(z)\} \quad \text{for } -\frac{\pi}{2} \leqslant z \leqslant \frac{\pi}{2}$$

and the Cauchy cumulative distribution function,

$$F(z) = \tfrac{1}{2} + \frac{1}{\pi} \tan^{-1}(z), \quad \text{for } -\infty < z < \infty$$

sometimes called Urban's curve. Appropriate graphs are given in Ashton (1972, p. 12). For most data sets it is impossible to distinguish between these different models (Cox and Snell, 1989, p. 22). An advantage of the logit model is its computational simplicity. It is also widely used in other areas, such as in the analysis of case-control data (Rosenbaum, 1991) and for projections of population size (Leach, 1981). For data sets containing as many individuals as Table 1.4, for example, it can prove possible to prefer, say, the probit model to the logit model. We return to this point in Chapter 4, where we also encounter a wide range of more complex models for quantal response data.

The binomial distribution was encountered in section 1.5, and it clearly forms the fundamental basis for much statistical analysis of quantal response data. We shall see later when it is replaced by alternatives such as Poisson and trinomial. A particularly important development is needed when 'over-dispersion', relative say to binomial or Poisson distributions, is present. In such cases variances are inflated, due to heterogeneity, relative to the simpler distributional forms. This is the topic of Chapter 6.

A common feature of the examples of section 1.3 is that the data are not obtained in a sequential manner. In some situations it is possible to employ a sequential design, and conduct new experiments at dose-levels suggested by responses to earlier experiments. This can result in a concentration of resources on the regions of doses of most overall interest, and we consider sequential methods as well as the general topic of how to design experiments in Chapter 8.

A fundamental feature of much bioassay, implicit in Examples 1.3, 1.4 and 1.9 amongst others, is the need for comparisons. This is a topic which recurs throughout the book, in section 3.7, and Chapters 6 and 7.

We have seen that analysis may be confounded by the doses which actually have an effect sometimes being different from those which nominally are administered. A further complication which may arise is if the substances tested are impure, or otherwise result in a mixture of substances being used. This is a subject which is dealt with at length in Plackett and Hewlett (1979). In some cases dose-response curves may be non-monotonic, in contrast to most of the examples above, and this may be due to the effect of mixtures of substances (e.g. Finney, 1971, p. 266), or possibly to high dose toxicity of substances which produce positive responses to lower doses (e.g.

Simpson and Margolin, 1986 and Exercise 1.16). These features, of response to mixtures, and of high-dose toxicity are considered respectively in section 3.7 and Chapter 4.

Note finally that much of the material that follows can be thought of as building blocks and these may be combined as appropriate for any particular problem, in ways which are not covered in this book.

The main texts in the area are those by: Aldrich and Nelson (1984), Ashton (1972), Cox and Snell (1989), Cramer (1991), Dobson (1990), Finney (1971, 1978), Govindarajulu (1988), Hubert (1992), Plackett and Hewlett (1979) and Salsburg (1986). Relevant material is also to be found in Agresti (1984), Bishop *et al.* (1975) and Haberman (1974). A useful case-study in logistic regression is given by Kay and Little (1986), and issue no. 4 of volume 42 of *Statistica Neerlandica* of 1988 is devoted to aspects of logistic regression.

The model-fitting required in the analysis of quantal response data nearly always involves numerical procedures for fitting non-linear models, and so inevitably computers are involved. In some cases, specialist computer packages and algorithms are available (e.g. Finney, 1976; Smith 1983; Russell *et al.* 1977; Morgan and Pack, 1988; Morgan *et al.* 1989). In more complex applications recourse is made to non-linear optimization routines (as in Rai and Van Ryzin, 1981, for example). The facilities of packages for generalized linear models, available for instance in GLIM, SAS and GENSTAT 5, are often particularly convenient. Also the MLP package (Ross, 1987) has routines for the analysis of quantal response data. At times throughout the book illustrations are made through computer code, with examples in GLIM, MINITAB and BASIC. A review of relevant programs and packages is given in Appendix E.

1.8 Exercises and complements

The exercises vary in complexity. It is particularly desirable to attempt the five exercises marked with a †. Exercises marked with a * are generally more difficult or speculative.

1.1 Identify at least three examples of quantal response data which do not arise from toxicology or efficacy studies.

1.2 Use a Taylor series expansion to verify the following first-order approximation, for sufficiently small σ and well-behaved function

$\psi(\ )$

$$V\{\psi(X)\} \approx \{\psi'(\mu)\}^2\sigma^2, \quad \text{where } E[X] = \mu,\ V(X) = \sigma^2$$

Deduce that if a random variable R has the binomial Bin(n, p) distribution, and is observed to be $R = r$, then we may estimate

$$V\left\{\log\left(\frac{R}{n-R}\right)\right\} \approx \frac{1}{r} + \frac{1}{(n-r)}$$

1.3† The following data sets were part of a collection put together by Copenhaver and Mielke (1977), and analysed also by Morgan (1985). Each set is given in the form: dose, dosage, concentration (or transformation of this), number responding, number in experiment. Details of dose transformations are given in Copenhaver and Mielke (1977, Table 1) and the raw data are given in Egger (1979), who excluded data set 14 on account of response being partly due to natural mortality. (note that this topic will be discussed in section 3.2.) The same numbering of data sets as in Copenhaver and Mielke (1977) has been adopted. In which cases do you think a preliminary log transformation of the dose scale has taken place? In each case, plot the proportions, and the logits of the proportions, and fit weighted and unweighted regressions on the logits. Comment on similarities and differences, with regard to estimates of error, as well as point estimates.

Set 2	2	10	30	Set 3	1	0	6
	3	14	30		2	0	6
	5	20	30		4	1	6
	9	23	30		8	0	6
					16	2	6
					32	4	6
					64	4	6
					128	6	6
					256	5	6
Set 4	0.41	6	50	Set 7	0	2	30
	0.58	16	48		1	8	30
	0.71	24	46		2	15	30
	0.89	42	49		3	23	30
	1.01	44	50		4	27	30

Set 9	1.6907	2	29	Set 10	0.71	16	49
	1.7242	7	30		1.0	18	48
	1.7552	9	28		1.31	34	48
	1.7842	14	27		1.48	47	49
	1.8113	23	30		1.61	47	50
	1.8369	29	31		1.7	48	48
	1.8610	29	30				
	1.8839	29	29				
Set 11	0.4	7	47	Set 14	1.57	34	132
	0.71	22	46		2.17	40	51
	1.0	27	46		2.49	114	127
	1.18	38	48		2.66	115	117
	1.31	43	46		2.79	125	125
	1.4	48	50				
Set 18	1	0	40	Set 22	0.4472	0	10
	2	2	40		0.55	2	10
	3	14	40		0.6232	3	10
	4	19	40		0.7	4	10
	5	30	40		0.7482	7	10
					0.8	8	10

Performing these calculations by hand can be quite time-consuming, even with the aid of special graph paper. Access to a computer package such as MINITAB can result in an enormous simplification. The following MINITAB commands may be used to answer this question for data in a file named: 'QUANTAL' and in the same format as above.

```
READ 'QUANTAL' INTO COLUMNS C1,C2,C3
LET C4=1/C3
LET C5=C2-C3
LET C6=1/C5
LET C7=C4+C6
LET C8=1/C7
NOTE: C8 CONTAINS THE WEIGHTS
LET C9=C3/C5
LET C10=LOG(C9)
NOTE: C10 CONTAINS THE LOGITS
PLOT C10 VS C1
REGRESS C10, 1 COVARIATE, ON C1
REGRESS C10, WEIGHTS IN C8, 1 COVARIATE, ON C1
```

Problems will be encountered with 0% or 100% responses: MINITAB overcomes these by failing to return a logit in that case,

and the above program can then still be run. A further collection of sets of quantal response data, in this case describing toxic response, is given in Rai and Van Ryzin (1979).

1.4 The data below, describing the mortality of adult flour-beetles (*Tribolium confusum*), after 5 hours' exposure to gaseous carbon disulphide (CS_2), are reproduced in Prentice (1976b).

Dose (CS_2 mg/l)	49.06	52.99	56.91	60.84	64.76	68.69	72.61	76.54
Number of beetles in experiment	59	60	62	56	63	59	62	60
Number of insects killed	6	13	18	28	52	53	61	60

Investigate whether a preliminary log transformation of the dose simplifies a plot of the logits. We shall discuss the analysis of these data further in Chapter 4. Taken from a paper by Bliss (1935), they form one of the standard data sets which are used to illustrate any new model for analysing quantal response data. This is probably due to their use as an illustration by Prentice (1976b), which was an early paper presenting new models.

1.5 The following data come from an experiment conducted at the Stanford School of Medicine to measure the toxicity of guthion on mice:

Dosage (mg/kg of body weight)	*Number of mice treated*	*Number of mice killed*
4	30	1
5	46	3
6	46	13
7	46	23
8	46	29
10	46	44

For these data, perform the same investigation as in Exercise 1.4.

1.6† A group of 16 pregnant female rats was fed a control diet during pregnancy and lactation. The diet of a second group of 16 pregnant females was treated with a chemical. For each litter the number n of pups alive at 4 days and the number r of pups that survived the 21-day lactation period were recorded. The resulting data were as follows:

Control	r	13	12	9	9	8	8	12	11	9	9	8	11	4	5	7	7
Group	n	13	12	9	9	8	8	13	12	10	10	9	13	5	7	10	10
Treated	r	12	11	10	9	10	9	9	8	8	4	7	4	5	3	3	0
Group	n	12	11	10	9	11	10	10	9	9	5	9	7	10	6	10	7

Provide a brief critical discussion of these data, and a simple test of whether the treatment affects survival. (Data from Weil, 1970, analysed in Williams, 1975.) We continue discussion of these data in section 6.2.

1.7 Kooijman (1981) presented the data below which are the cumulative mortality counts for *Daphnia magna* in water containing cadmium chloride. Plot the data and provide an analysis and a description of your conclusions. This is part of a larger data set, given in full in Table 5.1.

	Concentration (μg/l)					
Day	0.0	3.2	5.6	10.0	18.0	32.0
11	0	1	2	5	27	50
14	0	1	3	8	36	50
16	0	1	3	10	36	50
18	0	1	5	10	38	50
21	1	2	7	12	42	50
Group size	50	50	49	50	53	50

*1.8** Carter and Hubert (1981b, 1984) consider ways of analysing data of the following kind (taken from Carter and Hubert, 1981b) in which groups of 20 trout fry are exposed to different concentrations of copper sulphate. Observations were made over a period of 48 hours, and the experiment was repeated each week for 5 weeks. Presented are the cumulative numbers of dead fish. The stock of fish was homogeneous with respect to age at the start of the 5-week period, so that the ages of the fish used each week increased. At the