CHAPTER 8

Models for data with constant coefficient of variation

8.1 Introduction

The classical linear models introduced in Chapter 3 assume that the variance of the response is constant over the entire range of parameter values. This property is required to ensure that the regression parameters are estimated with maximum precision by ordinary least squares and that consistent estimates are obtained for the variance of $\hat{\boldsymbol{\beta}}$. It is common, however, to find data in the form of continuous measurements where the variance increases with the mean. In Chapter 6 we studied models for common types of data in which $\text{var}(Y) \propto E(Y)$, including continuous measurements as well as discrete data. Here we assume that the coefficient of variation is constant, i.e. that

$$\text{var}(Y) = \sigma^2\{E(Y)\}^2 = \sigma^2\mu^2.$$

Note that σ is now the coefficient of variation of Y and not the standard deviation.

For small σ, the variance-stabilizing transformation, $\log(Y)$, has approximate moments

$$E\big(\log(Y)\big) = \log(\mu) - \sigma^2/2 \qquad \text{and} \qquad \text{var}\big(\log(Y)\big) \simeq \sigma^2.$$

Further, if the systematic part of the model is multiplicative on the original scale, and hence additive on the log scale, then

$$\eta_i = \log\{E(Y_i)\} = \mathbf{x}_i^T\boldsymbol{\beta}.$$

Then, with the exception of the intercept or constant term in the linear model, consistent estimates of the parameters and of their precision may be obtained by transforming to the log scale and applying ordinary least squares. The intercept is then biased by approximately $-\sigma^2/2$.

For a number of reasons, and particularly if it is required to present conclusions on the original scale of measurement, it is preferable to retain that scale and not to transform the response. Then we have

$$\mu = E(Y) = \exp(\mathbf{x}^T\boldsymbol{\beta})$$

referring directly to the original scale of measurement. The log link function achieves linearity without abandoning the preferred scale, and a quadratic variance function describes the relationship between $\mathrm{var}(Y)$ and $E(Y)$. With this combination of link and variance function, iteratively weighted non-linear least squares may be used to obtain estimates for $\boldsymbol{\beta}$ using the algorithm described in Chapter 1. This method of estimation is equivalent to assuming that Y has the gamma distribution with constant index $\nu = 1/\sigma^2$ independent of the mean in the same sense that ordinary least squares arises as maximum likelihood for the Normal distribution.

In comparing the two methods of estimation described above, we assume that the precise distribution of Y is not specified, for if it is the comparison can be uniquely resolved by standard efficiency calculations and by considerations such as sufficiency. For example, if Y has the log-normal distribution the first method is preferred, while if it has the gamma distribution the second method is preferred. More generally, however, if Y is a variable with a physical dimension or if it is an extensive variable (Cox and Snell, 1981, p. 14) such that a sum of Ys has some well-defined physical meaning, the method of analysis based on transforming to $\log Y$ is unsatisfactory on scientific grounds and the second method of analysis would be preferred. However, if the analysis is exploratory or if only graphical presentation is required, transformation of the data is convenient and indeed desirable.

Firth (1988) gives a comparison of the efficiencies of the gamma model when the errors are in fact log-Normal with the log-Normal model when the errors have a gamma distribution. He concludes that the gamma model performs slightly better under reciprocal misspecification.

8.2 The gamma distribution

For our present purposes it is most convenient to write the gamma density in the form

$$\frac{1}{\Gamma(\nu)}\left(\frac{\nu y}{\mu}\right)^{\nu}\exp\left(-\frac{\nu y}{\mu}\right)d(\log y); \qquad y \geq 0, \nu > 0, \mu > 0.$$

For brevity we write $Y \sim G(\mu, \nu)$. From its cumulant generating function, $-\nu\log(1-\mu t/\nu)$, the first four cumulants are easily found as

$$\begin{aligned}\kappa_1 &= E(Y) = \mu,\\ \kappa_2 &= \mathrm{var}(Y) = \mu^2/\nu,\\ \kappa_3 &= E(Y-\mu)^3 = 2\mu^3/\nu^2,\\ \kappa_4 &= 6\mu^4/\nu^3.\end{aligned}$$

More generally $\kappa_r = (r-1)!\,\mu^r/\nu^{r-1}$. The value of ν determines the shape of the distribution. If $0 < \nu < 1$ the density has a pole at the origin and decreases monotonically as $y \to \infty$. The special case $\nu = 1$ corresponds to the exponential distribution. If $\nu > 1$ the density is zero at the origin and has a single mode at $y = \mu - \mu/\nu$; however, the density with respect to the differential element $d(\log y)$ has a maximum at $y = \mu$ for all ν. Fig. 8.1 shows the form of the distribution for $\nu = 0.5, 1.0, 2.0$ and 5.0 with $\mu = 1$ held constant. It can be seen from the graphs that the densities are all positively skewed. The standardized skewness coefficient is $\kappa_3/\kappa_2^{3/2} = 2\nu^{-1/2}$, and a Normal limit is attained as $\nu \to \infty$.

In this chapter we are concerned mostly with models for which the index or precision parameter $\nu = \sigma^{-2}$ is assumed constant for all observations, so that the densities all have the same shape. However, by analogy with weighted linear least squares, where the variances are proportional to known constants, we may, in the context of the gamma distribution, allow ν to vary in a similar manner from one observation to another. In other words we may have ν_i = constant × w_i, where w_i are known weights and ν_i is the index or precision parameter of Y_i. Problems of this form occur in estimating variance components where the observations are sums of squares of Normal variables, the weights are one half of their degrees of freedom and the proportionality constant is 1.

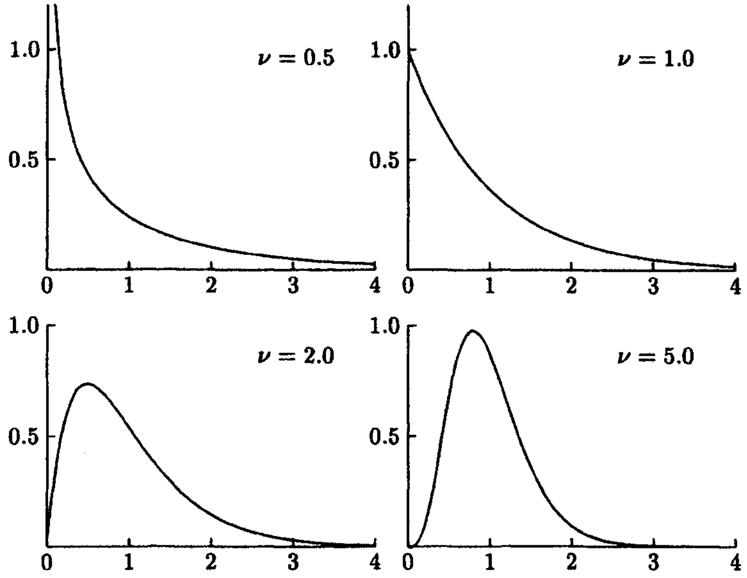

Fig. 8.1. *The gamma distribution for* $\nu = 0.5$, 1.0, 2.0 *and* 5.0, $\mu = 1$.

The densities then have different shapes. For further details see section 8.3.5.

The gamma family, and indeed all distributions of the type discussed in section 2.2, are closed under convolutions. Thus if $Y_1, \ldots, Y_n$ are independent and identically distributed in the gamma distribution with index ν, then the arithmetic mean $\bar{Y}$ is distributed in the same family with index $n\nu$. Thus the gamma distribution with integer index, sometimes also called the Erlangian distribution (Cox, 1962), arises in a fairly natural way as the time to the νth event in a Poisson process.

The log-likelihood function corresponding to a single observation is shown in Fig. 8.2 where we plot the log likelihood against μ, $\log \mu$, $\mu^{-1/3}$ and μ^{-1}. It can be seen that the log-likelihood function is nearly quadratic on the inverse cube-root scale; the log likelihood at μ differs from the value at the maximum by an amount closely approximated by

$$9y^{\frac{2}{3}}(y^{-\frac{1}{3}} - \mu^{-\frac{1}{3}})^2/2.$$

Now it is known that the square root of twice the log-likelihood-ratio statistic is approximately Normally distributed. Thus an

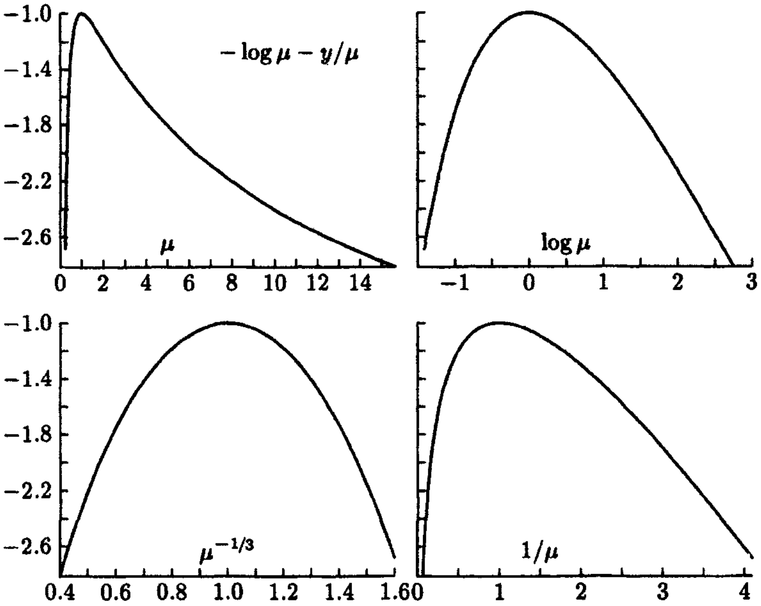

Fig. 8.2. *The gamma log likelihood for $y = 1$, plotted against μ, $\log\mu$, $\mu^{-1/3}$ and $1/\mu$.*

accurate Normalizing transformation for Y is

$$3\{(Y/\mu)^{\frac{1}{3}} - 1\}.$$

The cube-root transform was originally derived in this context by Wilson and Hilferty (1931) (see also Hougaard, 1982).

8.3 Models with gamma-distributed observations

8.3.1 *The variance function*

We have already noted that, with the parameterization of the gamma distribution used here, the variance function is quadratic. This result can be obtained directly by writing the log likelihood as a function of both ν and μ in the standard form

$$\nu(-y/\mu - \log\mu) + \nu\log y + \nu\log\nu - \log\Gamma(\nu).$$

It follows in terms of the parameterization used in Chapter 2 that $\theta = -1/\mu$ is the canonical parameter, and $b(\theta) = -\log(-\theta)$ is the cumulant function. From these the mean $b'(\theta) = \mu$ and variance function $b''(\theta) = \mu^2$ may be derived.

8.3.2 *The deviance*

Taking ν to be a known constant, the log likelihood may be written as

$$\sum_i \nu(-y_i/\mu_i - \log\mu_i)$$

for independent observations. If the index is not constant but is proportional to known weights, $\nu_i = \nu w_i$, the log likelihood is equal to

$$\nu \sum w_i(-y_i/\mu_i - \log\mu_i).$$

The maximum attainable log likelihood occurs at $\boldsymbol{\mu} = \boldsymbol{y}$, and the value attained is $-\nu\sum w_i(1 + \log y_i)$, which is finite unless $y_i = 0$ for some i. The deviance, which is proportional to twice the difference between the log likelihood achieved under the model and the maximum attainable value, is

$$D(\mathrm{y}; \hat{\boldsymbol{\mu}}) = -2\sum w_i\{\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)/\hat{\mu}_i\}.$$

This statistic is defined only if all the observations are strictly positive. More generally, if some components of $\mathbf{y}$ are zero we may replace $D(\mathbf{y}; \boldsymbol{\mu})$ by

$$D^+(\mathrm{y}; \hat{\boldsymbol{\mu}}) = 2C(\mathbf{y}) + 2\sum w_i \log\hat{\mu}_i + 2\sum w_i y_i/\hat{\mu}_i,$$

where $C(\mathbf{y})$ is an arbitrary bounded function of $\mathbf{y}$. The only advantage of $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ over $D^+(\mathbf{y}; \hat{\boldsymbol{\mu}})$ is that the former function is always positive and behaves like a residual sum of squares. Note, however, that the maximum-likelihood estimate of ν is a function of $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ and not of $D^+(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Furthermore, if any component of $\mathbf{y}$ is zero then $\hat{\nu} = 0$. This is clearly not a desirable feature of the maximum-likelihood estimator in most applications if only because rounding errors may produce spurious zeros. Alternative estimators are given in section 8.3.6.

The final term in the expression for $D(\mathbf{y};\hat{\boldsymbol{\mu}})$ is identically zero provided that the model formula contains an intercept term. In such cases the final term can be ignored (Nelder and Wedderburn, 1972). Under the same conditions, the final term in $D^{+}(\mathbf{y};\hat{\boldsymbol{\mu}})$ is equal to $\sum w_i$ and can be absorbed into $C(\mathbf{y})$.

8.3.3 *The canonical link*

The canonical link function yields sufficient statistics which are linear functions of the data and it is given by

$$\eta = \mu^{-1}.$$

Unlike the canonical links for the Poisson and binomial distributions, the reciprocal transformation, which is often interpretable as the rate of a process, does not map the range of μ onto the whole real line. Thus the requirement that $\eta > 0$ implies restrictions on the βs in any linear model. Suitable precautions must be taken in computing $\hat{\boldsymbol{\beta}}$ so that negative values of $\hat{\mu}$ are avoided.

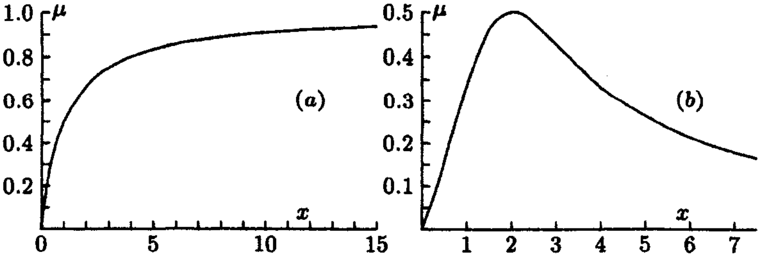

Fig. 8.3. *Inverse polynomials: (a) the inverse linear, $\mu^{-1} = 1 + x^{-1}$; (b) the inverse quadratic, $\mu^{-1} = x - 2 + 4/x$.*

An example of the canonical link is given by the inverse polynomial response surfaces discussed by Nelder (1966). The simplest case, that of the inverse linear response, is given by

$$\eta = \beta_0 + \beta_1/x \qquad \text{with} \quad x > 0.$$

In plant density experiments it is commonly observed that the yield per plant varies inversely with plant density x, so that the mean yield per plant has the form $1/(\beta_1 + \beta_0 x)$. The yield per unit area is then given by

$$\eta^{-1} = \mu = \frac{x}{\beta_0 x + \beta_1},$$

giving a hyperbolic form for μ against x, with a slope at the origin of $1/\beta_1$ and an asymptote at $\mu = 1/\beta_0$. Inclusion of a linear term in x, gives

$$\eta = \beta_1/x + \beta_0 + \gamma_1 x,$$

which is called the inverse quadratic. Both curves have a slope at the origin of $1/\beta_1$. The inverse quadratic response reaches a maximum of $\mu = \beta_0 + 2\sqrt{(\beta_1\gamma_1)}$ at $x = \sqrt{(\beta_1/\gamma_1)}$ corresponding to the optimum plant density. At higher plant densities μ tends to zero like $1/(\gamma_1 x)$ as shown in Fig. 8.3.

The surfaces can be extended to include more than one covariate and by the inclusion of cross-terms in $1/(x_1 x_2), x_1/x_2$, and so on. For positive values of the parameters the surfaces have the desirable property that the ordinate η is everywhere positive and bounded; this is in contrast to ordinary polynomials where the ordinate is unbounded at the extremes and often takes negative values.

In practice we often require to fit origins for the covariates, i.e. to make x enter the inverse polynomial in the form $x_0 + x$, where x_0 has to be estimated. The baseline value x_0 is non-linear in a general sense and its estimation requires special treatment—see Chapter 11 for details.

Two other link functions are important for generalized linear models with gamma errors, the log and the identity, and we now consider their uses.

8.3.4 *Multiplicative models: log link*

By combining the log link with terms linear in x and $1/x$ a large variety of qualitatively distinct response functions can be generated. Four of these are shown in Fig. 8.4, where we have shown $\eta = \log\mu = 1 \pm x \pm 1/x$ for $x > 0$. These curves are sometimes useful for describing response functions that have horizontal or vertical asymptotes, or functions that have turning points but are noticeably asymmetric about that point.

We noted in section 8.1 the close connection between linear models with constant variance for $\log Y$ and multiplicative models

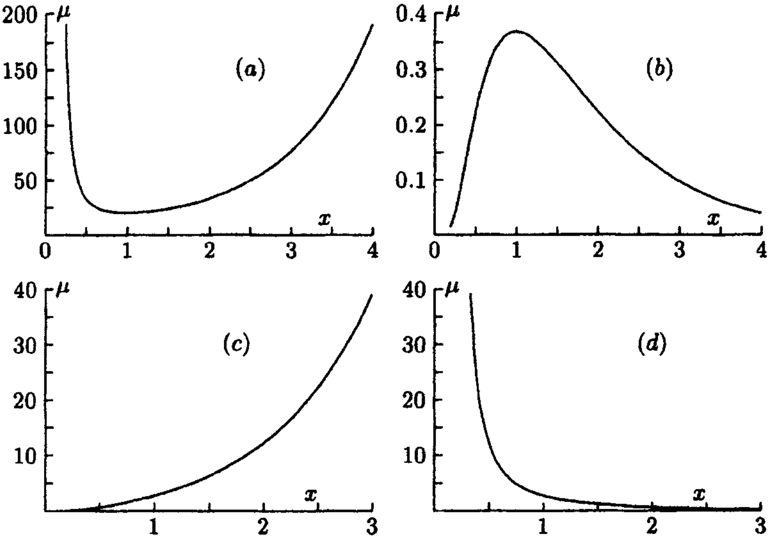

Fig. 8.4. *Plots of various logarithmic functions having asymptotes:*
(*a*) $\log(\mu) = 1 + x + 1/x$, (*b*) $\log(\mu) = 1 - x - 1/x$,
(*c*) $\log(\mu) = 1 + x - 1/x$, (*d*) $\log(\mu) = 1 - x + 1/x$.

with constant coefficient of variation for Y. Suppose that σ^2 is sufficiently small so that $\text{var}(\log Y) = \sigma^2 = \text{var}(Y)/\mu^2$. In a linear model for $\log Y$ the covariance matrix of the parameter estimates is $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, where $\mathbf{X}$ is the model matrix. For the corresponding multiplicative model the quadratic weight function is exactly unity, giving $\text{cov}(\hat{\boldsymbol{\beta}}) \simeq \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ as before. In particular, if $\mathbf{X}$ is the incidence matrix corresponding to an orthogonal design, so that parameter estimates in the Normal-theory linear model are independent, then the corresponding parameter estimates in the gamma-theory multiplicative model are independent asymptotically. This property of approximate independence holds for all generalized linear models whenever the link function is the same as the variance-stabilizing transform.

The preceding analysis and the discussion in section 8.1 indicate that for small σ^2 it is likely to be difficult to discriminate between Normal-theory linear models for $\log Y$ and gamma-theory multiplicative models for Y. Atkinson's (1982) work confirms this assertion even for σ^2 as large as 0.6.

8.3.5 *Linear models: identity link*

Sums of squares of independent Normal random variables have the chi-squared or, equivalently, the gamma distribution with known index w = (degrees of freedom)/2. One method of estimating variance components is to equate the observed mean squares y_i to their expectations which are linear functions of the unknown variance components. Thus

$$\mu_i = E(Y_i) = \sum x_{ij}\beta_j,$$

where x_{ij} are known coefficients and β_j are the variance components. Furthermore if the original data were Normally distributed,

$$\text{var}(Y_i) = \mu_i^2/w_i,$$

where w_i are known weights equal to one-half the degrees of freedom of Y_i. The preceding analysis can equally well be based on sums of squares rather than on mean squares; the coefficients x_{ij} would then be replaced by $2w_i x_{ij}$ and weights would be w_i because the coefficient of variation is unaffected by multiplication of the data by a constant.

If the number of variance components is the same as the number of mean squares, which is commonly the case, the estimating equations may be solved directly by inverting the set of linear equations. The method described above, based on the gamma likelihood, is required only when the number of independent mean squares exceeds the number of variance components. A further advantage of this procedure is that approximate asymptotic variances can be obtained for the estimated variance components. Unfortunately, the sizes of some of these variances often shows that the corresponding estimates are almost worthless. Normal-theory approximations for the distribution of $\hat{\beta}$s are usually very poor.

The analysis given above is based on the assumption that the mean-square variables are independent and that the original data were Normally distributed. Furthermore, negative estimates of variance components are not explicitly ruled out; these may, however, sometimes be interpretable (Nelder, 1977). In this respect weighted least squares is technically different from maximum likelihood, which does not permit negative variance components. If the weighted least-squares estimates turn out to be negative the likelihood function attains its maximum on the boundary of the parameter space corresponding to a zero variance component. The two methods coincide only if the weighted least-squares estimates are non-negative.

8.3.6 *Estimation of the dispersion parameter*

The approximate covariance matrix of the parameter estimates is $\text{cov}(\hat{\boldsymbol{\beta}}) \simeq \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$, where

$$\mathbf{W} = \text{diag}\{(d\mu_i/d\eta_i)^2/V(\mu_i)\}$$

is the $n \times n$ diagonal matrix of weights, $\mathbf{X}$ is the $n \times p$ model matrix and σ is the coefficient of variation. If σ^2 is known, the covariance matrix of $\hat{\boldsymbol{\beta}}$ may be computed directly; usually, however, it must be estimated from the residuals. Under the gamma model the maximum-likelihood estimate of $\nu = \sigma^{-2}$ is given by

$$2n\{\log\hat{\nu} - \psi(\hat{\nu})\} = D(\mathbf{y}; \hat{\boldsymbol{\mu}}), \tag{8.1}$$

where $\psi(\nu) = \Gamma'(\nu)/\Gamma(\nu)$. A suggested improvement to take account of the fact that p parameters have been estimated is to replace the l.h.s. of the above equation by

$$2n\{\log\hat{\nu} - \psi(\hat{\nu})\} - p\hat{\nu}^{-1}, \tag{8.2}$$

the correction being the $O(1)$ term in an asymptotic expansion for $E(D(\mathbf{Y}; \hat{\boldsymbol{\mu}}))$. There is a clear analogue here with the Normal-theory estimates of variance, $\hat{\sigma}^2$ and s^2. If ν is sufficiently large, implying σ^2 sufficiently small, we may expand (8.1) and (8.2) ignoring terms of order ν^{-2} or smaller. The maximum-likelihood estimate is then approximately

$$\hat{\nu}^{-1} \simeq \frac{\bar{D}(6+\bar{D})}{6+2\bar{D}}$$

where $\bar{D} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n$. A similar approximation can be made for the bias-corrected estimate: see Exercises 8.11 and 8.12. For further approximations, see the paper by Greenwood and Durand (1960) and a series of papers by Bain and Engelhardt (1975, 1977).

The principal problem with the maximum-likelihood estimator, and in fact with any estimator based on $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, is that it is extremely sensitive to rounding errors in very small observations and in fact $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ is infinite if any component of $\mathbf{y}$ is zero. Equally important is the fact that if the gamma assumption is false, $\hat{\nu}^{-1/2}$ does not consistently estimate the coefficient of variation. For these reasons we prefer the moment estimator

$$\tilde{\sigma}^2 = \sum\{(y - \hat{\mu})/\hat{\mu}\}^2/(n-p) = X^2/(n-p), \qquad (8.3)$$

which is consistent for σ^2, provided of course that $\boldsymbol{\beta}$ has been consistently estimated. This estimator for σ^2 may be used in the formula $\sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$ to obtain an estimate of $\text{cov}(\hat{\boldsymbol{\beta}})$. Note that, unlike the usual Normal-theory estimator of variance s^2, the bias of $\tilde{\sigma}^2$ is $O(n^{-1})$ even if the data are distributed according to the gamma distribution. The divisor $n - p$ is preferable to n but is not sufficient to remove the $O(n^{-1})$ bias. For a single gamma sample the expected value of $\tilde{\sigma}^2$ is

$$\sigma^2[1 - \sigma^2/n + O(n^{-2})].$$

The negative bias is a consequence of the fact that $V''(\mu) > 0$.

8.4 Examples

8.4.1 *Car insurance claims*

The data given in Table 8.1, taken from Baxter *et al.* (1980, Table 1), give the average claims for damage to the owner's car for privately owned and comprehensively insured vehicles. Averages are given in pounds sterling adjusted for inflation. The number of claims on which each average is based is given in parallel. Three factors thought likely to affect the average claim are:

1. policyholder's age (PA), with eight levels, 17–20, 21–24, 25–29, 30–34, 35–39, 40–49, 50–59, 60+;
2. car group (CG), with four levels, A, B, C and D;
3. vehicle age (VA), with four levels, 0–3, 4–7, 8–9, 10+.

The numbers of claims m_{ijk} on which each average is based vary widely from zero to a maximum of 434. Since the precision of each average Y_{ijk}, whether measured by the variance or by the squared coefficient of variation, is proportional to the corresponding m_{ijk}, these numbers appear as weights in the analysis. This means that the five accidentally empty cells, (1,3,4), (1,4,3), (1,4,4), (2,4,4) and (5,4,4) for which $m = 0$, are effectively left out of the analysis. For computational purposes, however, it is usually more convenient to retain these cells as observations with zero weight, so that they make no contribution to the likelihood. The structure of the factors is then formally that of a complete crossed design.

Baxter *et al.* analyse the data using a weighted Normal-theory linear model with weights m_{ijk} and the three main effects PA + CG + VA. Here we reanalyse the data, making the assumption that the coefficient of variation rather than the variance is constant across cells. In addition, we make the assumption that the systematic effects are linear on the reciprocal scale rather than on the untransformed scale. Justification for these choices is given in Chapters 10 and 11. The model containing main effects only may be written

$$\mu_{ijk} = E(Y_{ijk}) = (\mu_0 + \alpha_i + \beta_j + \gamma_k)^{-1},$$
$$\operatorname{var}(Y_{ijk}) = \sigma^2 \mu_{ijk}^2 / m_{ijk},$$

where α_i, β_j and γ_k are the parameters corresponding to the three classifying factors PA, CG and VA. One way of interpreting the reciprocal transform is to think of $\eta_{ijk} = 1/\mu_{ijk}$ as the rate at which instalments of £1 must be paid to service an average claim over a fixed period of one time unit. In other words, η_{ijk} is the time interval between instalments or the time purchased by an instalment of £1 in servicing an average claim in cell (i, j, k).

One sequence of models yielded the goodness-of-fit statistics shown in Table 8.2. Using the result that first differences of the deviance have, under the appropriate hypothesis, an approximate scaled chi-squared distribution, it is clear that the model with main effects only provides a reasonable fit and that the addition of two-factor interactions yields no further explanatory power. The estimate of $\tilde{\sigma}^2$ based on the residuals from the main-effects model is

$$\tilde{\sigma}^2 = \frac{1}{109} \sum m(y - \hat{\mu})^2 / \hat{\mu}^2 = 1.21,$$

so that the estimated coefficient of variation of the individual claims is $\tilde{\sigma} = 1.1$. Estimates based on the deviance give very similar values. An examination of approximately standardized residuals

Table 8.1 *Average cost of claims for own damage (adjusted for inflation) for privately owned, comprehensively insured cars in* 1975

Policy-holder's age	*Car group*	*Vehicle age*							
		0–3		4–7		8–9		10+	
		£	No.	£	No.	£	No.	£	No.
17–20	A	289	8	282	8	133	4	160	1
	B	372	10	249	28	288	1	11	1
	C	189	9	288	13	179	1	—	0
	D	763	3	850	2	—	0	—	0
21–24	A	302	18	194	31	135	10	166	4
	B	420	59	243	96	196	13	135	3
	C	268	44	343	39	293	7	104	2
	D	407	24	320	18	205	2	—	0
25–29	A	268	56	285	55	181	17	110	12
	B	275	125	243	172	179	36	264	10
	C	334	163	274	129	208	18	150	8
	D	383	72	305	50	116	6	636	1
30–34	A	236	43	270	53	160	15	110	12
	B	259	179	226	211	161	39	107	19
	C	340	197	260	125	189	30	104	9
	D	400	104	349	55	147	8	65	2
35–39	A	207	43	129	73	157	21	113	14
	B	208	191	214	219	149	46	137	23
	C	251	210	232	131	204	32	141	8
	D	233	119	325	43	207	4	—	0
40–49	A	254	90	213	98	149	35	98	22
	B	218	380	209	434	172	97	110	59
	C	239	401	250	253	174	50	129	15
	D	387	199	299	88	325	8	137	9
50–59	A	251	69	227	120	172	42	98	35
	B	196	366	229	353	164	95	132	45
	C	268	310	250	148	175	33	152	13
	D	391	105	228	46	346	10	167	1
60+	A	264	64	198	100	167	43	114	53
	B	224	228	193	233	178	73	101	44
	C	269	183	258	103	227	20	119	6
	D	385	62	324	22	192	6	123	6

Table 8.2 *Goodness-of-fit statistics for a sequence of models fitted to the car insurance data (error gamma; link reciprocal)*

Model	*Deviance*	*First difference*	*d.f.*	*Mean deviance*
1	649.9			
		82.2	7	11.7
PA	567.7			
		228.3	3	76.1
PA + CG	339.4			
		214.7	3	71.6
PA + CG + VA	124.8			
		34.0	21	1.62
+ PA·CG	90.7			
		19.7	21	0.94
+ PA·VA	71.0			
		5.4	9	0.60
+ CG·VA	65.6			
		65.6	58	1.13
Complete	0.0			

Table 8.3 *Parameter estimates and standard errors* ($\times 10^6$) *on reciprocal scale for main effects in car insurance example*

Level	*Age group* (PA)	*Car group* (CG)	*Vehicle age* (VA)
1	0 (—)	0 (—)	0 (—)
2	101 (436)	38 (169)	336 (101)
3	350 (412)	−614 (170)	1651 (227)
4	462 (410)	−1421 (181)	4154 (442)
5	1370 (419)		
6	970 (405)		
7	916 (408)		
8	920 (416)		

using the formula $\sqrt{m}(y-\hat{\mu})/(\tilde{\sigma}\hat{\mu})$ shows the six most extreme residuals as corresponding to observations (2,2,1), (3,2,4), (3,4,4), (5,1,2), (5,4,1) and (7,2,1) with values 3.4, 3.2, 2.8, −2.6, −2.2 and −2.5. The corresponding standardized deviance residuals are 3.0, 2.4, 1.8, −3.0, −2.3 and −2.7. The positions of these cells do not show any obvious pattern, and the magnitudes of the most extreme residuals are only moderately large in view of the sample size, which is effectively 109. With a Normal sample of this size one expects the most extreme standardized residuals to be about ±2.5.

Parameter estimates for the main-effects model are given in

Table 8.3. Standard errors are based on the estimate $\tilde{\sigma} = 1.1$. The estimate for the intercept corresponding to all factors at their lowest level is 3410×10^{-6}. Bearing in mind that the analysis is performed here on the reciprocal scale and that a large positive parameter corresponds to a small claim, we may deduce the following. The largest average claims are made by policyholders in the youngest four age groups, i.e. up to age 34, the smallest average claims by those aged 35–39, and intermediate claims by those aged 40 and over. These effects are in addition to effects due to type of vehicle and vehicle age. The value of claims decreases with car age, although not linearly. There are also marked differences between the four car groups, group D being the most expensive and group C intermediate. No significant difference is discernible between car groups A and B.

It should be pointed out that the parameter estimates given here are contrasts with level 1. In a balanced design the three sets of estimates corresponding to the three factors would be uncorrelated while the correlations within a factor would be 0.5. Even where, as here, there is considerable lack of balance, the correlations do not deviate markedly from these values.

It is possible to test and quantify the assertions made above by fusing levels 1–4, levels 6–8 of PA and levels 1 and 2 of CG. The deviance then increases to 129.8 on 116 d.f., which is a statistically insignificant increase.

The preceding analysis is not the only one possible for these data. In fact a multiplicative model corresponding to a logarithmic link function would lead to similar qualitative conclusions. As is shown in Chapter 10, the data themselves support the reciprocal model better but only marginally so, and it might be argued that quantitative conclusions for these data would be more readily stated and understood for a multiplicative model.

8.4.2 *Clotting times of blood*

Hurn *et al.* (1945) published data on the clotting time of blood, giving clotting times in seconds (y) for normal plasma diluted to nine different percentage concentrations with prothrombin-free plasma (u); clotting was induced by two lots of thromboplastin. The data are shown in Table 8.4. A hyperbolic model for lot 1 was fitted by Bliss (1970), using an inverse transformation of the data, and for both lots 1 and 2 using untransformed data. We analyse both lots using the inverse link and gamma errors.

Initial plots suggest a log scale for u to produce inverse linearity, with different intercepts and slopes for the two lots. This claim is confirmed by fitting the following model sequence:

Model	*Deviance*	*d.f.*
1	7.709	17
X	1.018	16
$L+X$	0.300	15
$L+L.X$	0.0294	14

Here $x = \log u$ and L is the factor defining the lots. Clearly all the terms are necessary and the model produces a mean deviance whose square root is 0.0458, implying approximately a 4.6% standard error on the y-scale. A plot of residuals against x shows that the fit is unsatisfactory for the two lots with $u = 5$. If these points are omitted from the fit, the residual standard error is reduced to 2.5%. The fitted lines, with nominal standard errors in parentheses, are

$$\begin{aligned}\text{lot 1:}\quad & \hat\mu_1^{-1} = -0.02177(0.00116) + 0.01691(0.00038)x,\\ \text{lot 2:}\quad & \hat\mu_2^{-1} = -0.02965(0.00186) + 0.02531(0.00061)x.\end{aligned}$$

The fitted values for $u = 5$ are then 183.5 and 90.1, which are about 50% larger than the observed values.

However, the fitted value given by the inverse linear model is unusually sensitive to recording errors or rounding errors in the value of u at the lower end of the range. If the lowest value is changed from $u = 5$ to $u = 5.5$ or $u = 6$ the parameter estimates are very close to those shown above, and the residual standard deviation is again estimated at 2.5%. In other words, the observed values do not appear to be consistent with the recorded concentration $u = 5$, but they are entirely consistent with $u = 6$.

The estimates suggest that the parameters for lot 2 are a constant multiple (about 1.6) of those for lot 1. If true this would mean that $\mu_2 = k\mu_1$, where the suffix denotes the lot. This model, though not a generalized linear model, has simple maximum-likelihood equations for estimating α, β and k where

$$\begin{aligned}\mu_1 &= 1/\eta_1, \qquad \eta_1 = \alpha + \beta \mathbf{x},\\ \mu_2 &= k\mu_1.\end{aligned}$$

Table 8.4 *Mean clotting times in seconds* (y) *of blood for nine percentage concentrations of normal plasma* (u) *and two lots of clotting agent*

	Clotting time	
u	*Lot* 1	*Lot* 2
5	118	69
10	58	35
15	42	26
20	35	21
30	27	18
40	25	16
60	21	13
80	19	12
100	18	12

These are equivalent to fitting α and β to data $\mathbf{y}_1$ and $\mathbf{y}_2/k$, combined with the equation $\sum(y_2/\mu_1 - k) = 0$. The resulting fit gives $\hat{k} = 0.625$ with deviance $= 0.0332$ and having 15 d.f. Comparing this with the fit of separate lines gives a difference of deviance of 0.0038 on one degree of freedom against a mean deviance of 0.0021 for the more complex model. The simpler model of proportionality is not discounted, with lot 2 giving times about five-eighths those of lot 1.

8.4.3 *Modelling rainfall data using two generalized linear models*

Histograms of daily rainfall data are usually skewed to the right with a 'spike' at the origin. This form of distribution suggests that such data might be modelled in two stages, one stage being concerned with the pattern of occurrence of wet and dry days, and the other with the amount of rain falling on wet days. The first stage involves discrete data and can often be modelled by a stochastic process in which the probability of rain on day t depends on the history of the process up to day $t-1$. Often, first-order dependence corresponding to a Markov chain provides a satisfactory model. In the second stage we require a family of densities on the positive line for the quantity of rainfall. To be realistic, this family of densities should be positively skewed and should have variance increasing with μ. The gamma distribution has been found appropriate in this context, although the log-Normal distribution is also widely used.

(a) *Modelling the frequency of wet days.* Coe and Stern (1982, 1984) describe the application of generalized linear models and give references to earlier work. The data for n years form an $n \times 365$ table of rainfall amounts. (We ignore the complications introduced by leap years.) Considering the years as replicates, each of the n observations of day t is classified by the double dichotomy dry/wet and previous day dry/previous day wet. Combining over replicates we obtain, for each of the 365 days, a 2×2 table of frequencies having the form of Table 8.5.

Table 8.5 *The* 2×2 *table of frequencies for rainfall data on day t*

		Today		
		Wet	*Dry*	*Total*
Yesterday	*Wet*	y_0	$n_0 - y_0$	n_0
	Dry	y_1	$n_1 - y_1$	n_1
	Total	$y_.$	$n_. - y_.$	$n_. = n$

Let $\pi_0(t)$ be the probability that day t is wet given that day $t-1$ was wet: $\pi_1(t)$ is the corresponding probability given that day $t-1$ was dry. Ignoring end effects, the likelihood for the first-order Markov model is the product over t of terms having the form

$$\pi_0(t)^{y_0}[1-\pi_0(t)]^{n_0-y_0}\pi_1(t)^{y_1}[1-\pi_1(t)]^{n_1-y_1}.$$

In other words each 2×2 table corresponds to two independent binomial observations in which the row totals are regarded as fixed.

Note that in the above 2×2 table for day t, n_0 is the number of occasions on which rain fell on day $t-1$ in the years $1, \ldots, n$, whereas $y_.$ is the number of occasions on which rain fell on day t. Evidently therefore, in an obvious extension of the notation, $n_0(t+1) = y_.(t)$.

If the target parameter were the difference between $\pi_0(t)$ and $\pi_1(t)$, say

$$\psi(t) = \pi_0(t)[1-\pi_1(t)]/\{\pi_1(t)(1-\pi_0(t))\},$$

it would often be preferable to construct a likelihood function depending on $\psi(t)$ alone. The hypergeometric likelihood described in section 7.4 can then be used. In this application, however, we would usually be interested in models for $\pi_0(t)$ and $\pi_1(t)$ themselves and not just in the difference between them.

Coe and Stern use linear logistic models with various explanatory terms. Obvious choices for cyclical terms are the harmonics $\sin(2\pi t/365)$, $\cos(2\pi t/365)$, $\sin(4\pi t/365)$, $\cos(4\pi t/365)$, and so on. The simplest model corresponding to the first harmonic is

$$\begin{aligned}\operatorname{logit}(\pi_0(t)) &= \alpha_0 + \alpha_{01}\sin(2\pi t/365) + \beta_{01}\cos(2\pi t/365),\\ \operatorname{logit}(\pi_1(t)) &= \alpha_1 + \alpha_{11}\sin(2\pi t/365) + \beta_{11}\cos(2\pi t/365),\end{aligned}$$

which involves six parameters. Note that if the coefficients of the harmonic terms are equal ($\alpha_{01} = \alpha_{11}, \beta_{01} = \beta_{11}$), the odds ratio in favour of wet days is constant over t.

If there is a well defined dry season, a different scale corresponding to some fraction of the year might be more appropriate.

Various extensions of these models are possible: a second-order model would take account of the state of the two previous days, producing four probabilities to be modelled. If it is suspected that secular trends over the years are present it is important not to regard the years as replicates. Instead, we would regard the data as $365n$ Bernoulli observations indexed by day, year and previous day wet/dry. The computational burden is increased but no new theoretical problems are involved.

(b) *Modelling the rainfall on wet days.* Coe and Stern use a multiplicative model with gamma-distributed observations to model the rainfall on wet days. The idea is to express $\log[\mu(t)]$ as a linear function involving harmonic components. Here $\mu(t)$ is the mean rainfall on day t conditional on day t being wet. If it is assumed that no secular trends are involved, the analysis requires only the means for each day of the period with the sample sizes entering the analysis as weights. The introduction of secular trends involves the use of individual daily values in the analysis. The assumption of constant coefficient of variation requires checking The simplest way is to group the data into intervals based on the value of $\hat{\mu}$ and to estimate the coefficient of variation in each interval. Plots against $\hat{\mu}$ should reveal any systematic departure from constancy.

(c) *Some results.* Coe and Stern present the results of fitting the models described above to data from Zinder in Niger spanning 30 years. The rainy season lasts about four months so that data are restricted to 120 days of the year. Table 8.6 shows the results of fitting separate Fourier series to $\pi_0(t)$ and $\pi_1(t)$ in a first-order Markov chain. Each new term adds four parameters to the model, a sine and cosine term for each π.

Table 8.6 *Analysis of deviance for a first-order Markov chain. Rainy days in rainfall data from Niger*

Model	*Deviance*	*First difference*	*d.f.*	*Mean deviance*
Intercept	483.1	—	238	
+1st harmonic	260.9	222.2	4	55.6
+2nd harmonic	235.6	25.3	4	6.3
+3rd harmonic	231.4	4.2	4	1.05
+4th harmonic	227.7	3.7	4	0.9

By including a sufficient number of harmonic terms in the model, the mean deviance is reduced to a value close to unity, indicating a satisfactory fit. The reductions for the third and fourth harmonic are clearly insignificant. Thus a first-order non-stationary Markov chain, with two harmonic terms for the time-dependence of the transition probabilities, is adequate for these data. This model contains a total of 10 parameters for the transition probabilities.

Table 8.7 *Analysis of deviance of rainfall amounts. Data from Niger*

Model	*Deviance*	*d.f.*	*Mean deviance*
Constant	224.6	119	
+1st harmonic	154.5	117	
+2nd harmonic	147.0	115	1.28
Within days	1205.0	946	1.27

The results of fitting models with gamma errors and log link for the rainfall amounts are shown in Table 8.7. Again two harmonics suffice in the sense that their inclusion reduces the between-day deviance to that within days. The mean deviance within days over years constitutes a baseline for the analysis between days, and its size (little more than 1) indicates a distribution of rainfall amounts on wet days that is close to exponential.

This analysis is based on the simplifying assumption that the probability of rain occurring on day t depends only on whether rain fell on day $t-1$, but not otherwise on the amount of rain. The decomposition into two independent generalized linear models depends heavily on this assumption. It is at least plausible, however, that the occurrence of rain on day t depends on whether or not it rained heavily on the previous day. Dependence of this nature can be checked by including in the logistic model for $\pi_0(t)$ the amount of rainfall (log units) on the previous day.

8.4.4 *Developmental rate of Drosophila melanogaster*

The data shown in Table 8.8 were collected by Powsner (1935) as part of an experiment to determine accurately the effect of temperature on the duration of the developmental stages of the fruit fly *Drosophila melanogaster*. Powsner studied four stages in its development, namely the embryonic, egg-larval, larval and pupal stages: only the first of these is considered here.

In all cases the eggs were laid at approximately 25°C and remained at that temperature for 20–30 minutes as indicated in the final column. Subsequently the eggs were brought to the experimental temperature, which was kept constant over the period of the experiment. Column 3 gives the average duration of the embryonic period measured from the time at which the eggs were laid. The number of eggs in each batch, together with the sample standard deviation of each batch, are shown in columns 4 and 5.

Figure 8.5 shows the batch standard deviations plotted against the batch means. Evidently, with the exception of the point at 32°C, the standard deviations are roughly proportional to the mean. A more formal weighted log-linear regression of the sample variances on the log of the sample means, with weights equal to the degrees of freedom, gives the fitted equation

$$\log(\text{sample variance}) \simeq -9.29 + 2.58 \log(\text{sample mean}),$$

suggesting a power relationship for the variance function with index in the range 2–3. In what follows, we assume that $V(\mu) = \mu^2$, implying that the coefficient of variation is constant. In other

Table 8.8 *Mean duration of embryonic period in the development of* Drosophila melanogaster

					Eggs laid at	
Temp. °C	*Exp. No.*	*Duration (hours)*	*Batch size*	*Std. dev.*	*Temp* °C	*Duration (hours)*
14.95	25	67.5 ± 0.33	54	2.41	25.1	0.33
16.16	44	57.1 ± 0.12	182	2.28	25.0	0.50
16.19	26	56.0 ± 0.12	153	1.46	25.1	0.33
17.15	28	48.4 ± 0.12	129	1.40	25.1	0.50
18.20	25	41.2 ± 0.16	64	1.30	25.1	0.33
19.08	33	37.80 ± 0.059	94	0.57	25.1	0.50
20.07	28	33.33 ± 0.080	82	0.73	25.1	0.33
22.14	25	26.50 ± 0.083	57	0.63	25.1	0.33
23.27	28	24.24 ± 0.038	135	0.44	25.1	0.50
24.09	33	22.44 ± 0.029	188	0.40	25.1	0.50
24.81	42	21.13 ± 0.017	217	0.36	25.0	0.50
24.84	40	21.05 ± 0.027	141	0.46	25.0	0.50
25.06	27	20.39 ± 0.064	37	0.38	25.1	0.50
25.06	27	20.41 ± 0.037	84	0.34	25.1	0.50
25.80	26	19.45 ± 0.026	196	0.36	25.1	0.33
26.92	33	18.77 ± 0.029	104	0.30	25.1	0.50
27.68	26	17.79 ± 0.041	148	0.49	25.1	0.33
28.89	29	17.38 ± 0.043	83	0.39	25.1	0.25
28.96	40	17.26 ± 0.031	95	0.43	25.0	0.50
29.00	44	17.18 ± 0.023	232	0.50	25.0	0.50
30.05	26	16.81 ± 0.032	148	0.39	25.1	0.33
30.80	26	16.97 ± 0.028	195	0.39	25.1	0.33
32.00	33	18.20 ± 0.290	58	2.23	25.1	0.50

Source: Powsner (1935).

words, the squared coefficient of variation of the batch means is assumed to be inversely proportional to the batch size. A similar analysis gives 0.0267 as a combined estimate of the coefficient of variation of the individual egg durations, ignoring the batch at 32°C.

The greatly increased variance for the batch of eggs maintained at 32°C suggests either that the tight experimental control was relaxed for this batch or, more plausibly, that the biochemistry of development at such an elevated temperature differs in important ways from the biochemistry at lower temperatures. One possibility is that the smaller eggs may suffer stress from dehydration at such

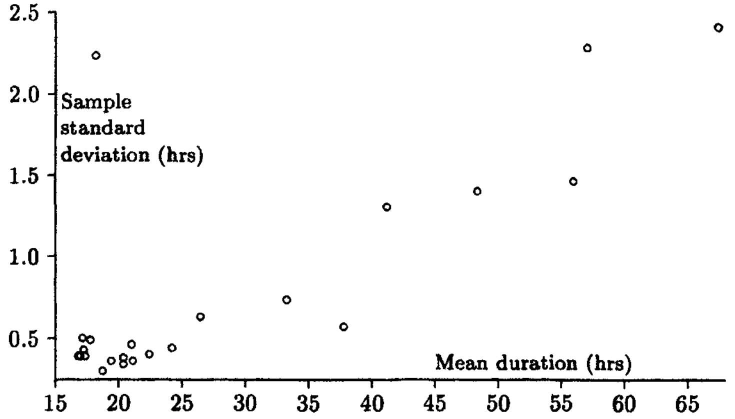

Fig. 8.5. *Plot of standard deviations against sample means for* 23 *batches of eggs. The outlying point corresponds to the highest temperature.*

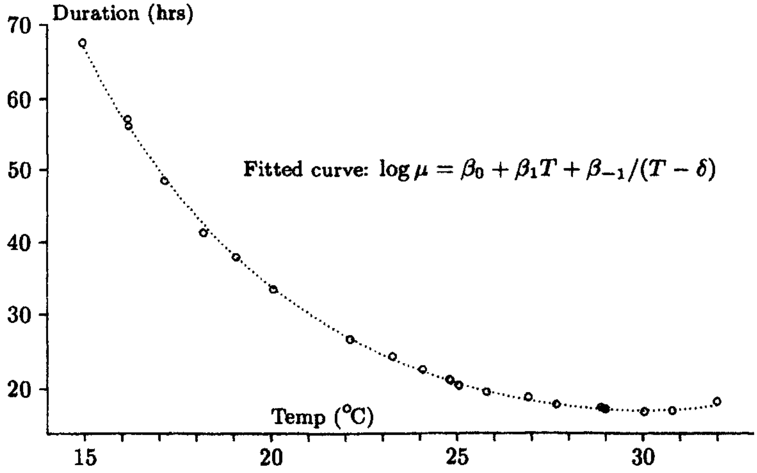

Fig. 8.6. *Observed average duration of embryonic period plotted against temperature (circles). The curve was fitted using gamma errors, log link and weighted by sample size.*

temperatures.

Figure 8.6 shows the observed mean durations plotted against temperature. Evidently the observed points lie very close to a smooth curve, which may have a minimum at around 29–31°C. To a large extent the evidence for a minimum rather than an asymptote rests on the observation at 32°C. However, it seems clear on general grounds that if the temperature is sufficiently high the eggs must begin to suffer, so that an eventual increase in duration is to be expected.

One of Powsner's objectives was to test whether the rates of complex biochemical reactions obey the laws that are known to govern simple chemical reactions. Arrhenius's law for the rate of simple chemical reactions is

$$\log \text{rate} = -\mu/(RT),$$

where μ is the 'critical chemical increment' for the reaction, R is the gas constant, and T is the absolute temperature. Since the reaction rate is inversely proportional to its duration, the Arrhenius model predicts a linear relationship in the graph of log(duration) against the reciprocal of absolute temperature. The observed graph for these data, however, is manifestly non-linear even when the point at 32°C is excluded. Consequently the simple Arrhenius model is unsatisfactory, even over small temperature ranges, as a description of the rate of development of *Drosophila melanogaster.*

Since the Arrhenius model is not even a close approximation to what is observed experimentally, it is necessary to proceed empirically without the support of tested theories. In what follows we treat the observed duration as the response having a squared coefficient of variation inversely proportional to the batch size. In all regression equations, therefore, each batch mean is weighted according to the batch size. This choice of response is not the only possibility: we could, for example, work with the observed duration minus the duration of the egg-laying period. The latter adjustment is rather small and is likely to have a very minor effect on the conclusions.

It is possible to obtain a reasonably good fit to the observed data by using cubic or quartic polynomials for the logarithmic reaction rate, i.e. using gamma errors with log link. However it seems worthwhile in this example to consider functions of temperature that have asymptotes. It is known for example that no development takes place below a certain critical temperature. There is undoubtedly a corresponding upper limit. Thus we are led to consider rational functions of temperature, the simplest of which is

$$\beta_0 + \beta_1 T + \beta_{-1}/(T - \delta). \qquad (8.4)$$

This can be expressed as a rational function in T whose denominator is linear and numerator quadratic. It is immaterial in (8.4) whether T is measured in °C, °F or °K.

So far we have not stated whether (8.4) is to be considered as a model for the rate, the log rate, or the duration of the embryonic period. These choices amount to different choices of link functions, namely the reciprocal, logarithm and identity respectively. The model is linear in $\boldsymbol{\beta}$ for each fixed δ. For a simple comparison among the three link functions, therefore, we take $\delta = 0$. The deviances are 5.97 for the reciprocal, 2.77 for the logarithm and 0.473 for the identity. Among these three choices, the identity link is strongly preferred and the fit is surprisingly good. It is visually indistinguishable from the curve plotted in Fig. 8.6.

Choosing $\delta = 0$ amounts to stating a preference for the Celsius scale over the Kelvin and Fahrenheit scales. Treating δ as a free parameter leaves the choice of scale in the hands of the data. The best-fitting linear model has $\hat{\delta} \simeq 0.6$°C. The best-fitting log-linear model has $\hat{\delta} \simeq 58.6$°C, as can be seen from Fig. 8.8a, while the best-fitting inverse-linear model has $\hat{\delta} \simeq 33.5$°C. The corresponding deviances are 0.47, 0.32 and 1.41 respectively. The log-rational fitted curve is shown in Fig. 8.6. Parameter estimates and nominal standard errors are shown in the table below.

Parameter estimates in the log-rational model (8.4)

Parameter	*Estimate*	s.e.
β_0	3.201	1.594
β_1	−0.265	0.0355
β_{-1}	−217.08	125.21
δ	58.644	6.48

The estimated minimum duration or maximum rate of embryonic development occurs at

$$\hat{T} = \hat{\delta} - (\hat{\beta}_{-1}/\hat{\beta}_1)^{1/2} = 30.01\text{°C}.$$

The residual coefficient of variation is estimated as

$$\tilde{\sigma} = (0.32/19)^{1/2} = (0.0168)^{1/2} = 0.13,$$

or 13%. This is the estimated coefficient of variation for the duration of the embryonic period of individual eggs. The estimated coefficient of variation of the batch means is then $0.13/\sqrt{m_i}$, where m_i is the batch size. Despite the exceptionally good fit obtained using this class of functions, the between-batch residual variation, as measured by the coefficient of variation, is substantially larger than the within-batches coefficient of variation. From columns 3 and 5 in Table 8.8, the within-batches coefficient of variation of the individual egg durations is estimated as approximately 2.7%. Thus the ratio of the between- to within-batch squared coefficient of variations is about 23:1. If a reasonable allowance were made for model selection, this ratio would be even larger.

It would appear, therefore, that apart from the temperature differences there must have been other unrecorded differences in experimental conditions from batch to batch, for example differences in humidity, lighting conditions, temperature ranges, ventilation and so on.

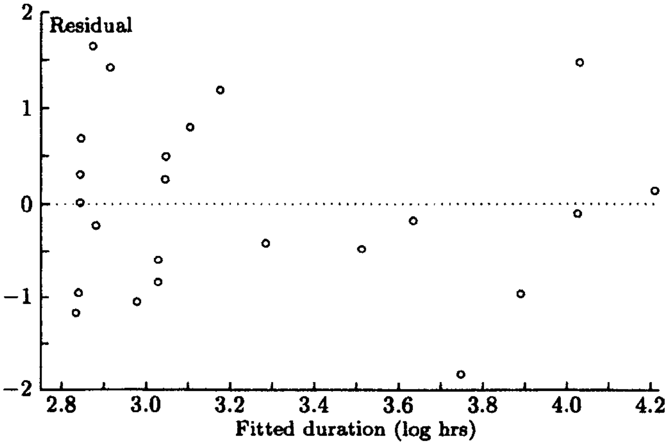

Fig. 8.7. *Plot of deviance residuals for model* (8.4) *against log fitted values for* 23 *batches of eggs.*

So far as the gamma assumption is concerned it is the constancy of the between-batches coefficient of variation and not the within-batch coefficient of variation that is relevant for model-fitting purposes. From that point of view, the diagram in Fig. 8.5 is irrelevant in deciding whether the variance function is quadratic. In order to test whether the between-batches coefficient of variation is constant, we examine the plot of the standardized deviance residuals plotted against fitted values (Fig. 8.7). These deviance residuals, including the weights, are given by

$$\pm\left[-2w_i\{\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)/\hat{\mu}_i\}\right]^{1/2}\Big/\tilde{\sigma},$$

where $\tilde{\sigma} = 0.13$ is the estimated between-batches coefficient of variation.

If the between-batches variance function is indeed quadratic, this residual plot should look like an ordinary Normal-theory residual plot. In fact all five residuals corresponding to fitted values in the range 25–50 are small but negative, so there is evidence of a small but systematic departure from the fitted model. However there is no evidence that the dispersion of the residuals increases or decreases systematically with the fitted values. Thus the between-batches coefficient of variation appears to be constant or nearly so.

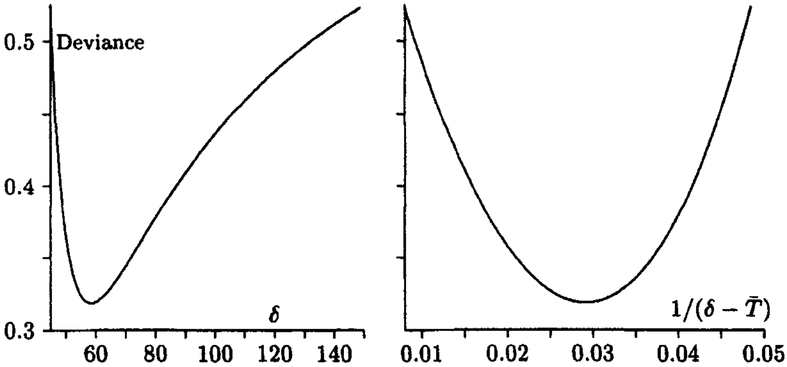

Fig. 8.8. *The deviance function for model* (8.4) *plotted against* δ *and against* $1/(\delta - \bar{T})$.

An awkward aspect of intrinsically non-linear models such as (8.4) is that Normal-theory approximations for the distribution of maximum-likelihood estimators may not be very accurate unless care is taken to make an appropriate transformation of the parameter. In particular, the distribution of $\hat{\delta}$ in (8.4) is noticeably non-Normal and accurate confidence intervals based on the deviance function are noticeably asymmetric. Figure 8.8a shows the residual deviance plotted against δ for values of δ in the range

45–150. Figure 8.8b shows the residual deviance plotted against $\zeta = 1/(\delta - \bar{T})$ over the equivalent range. Evidently, likelihood-based confidence limits for ζ are nearly symmetrically located about $\hat{\zeta}$, so that Normal-theory approximations for $\hat{\zeta}$ are more accurate than Normal-theory approximations for δ. Note that the transformation $\delta \rightarrow \zeta$ takes $\delta = \pm\infty$ to $\zeta = 0$. The likelihood function is continuous in ζ at this point.

For more complicated non-linear models such simplifying parameter transformations may not be easy to construct. In such cases it is necessary to plot the deviance function or log-likelihood function in order to obtain reasonably accurate confidence intervals.

The foregoing discussion presupposes the correctness of the assumed model. In this example, however an equally good fit to the observed points can be obtained using a polynomial model of degree four in place of (8.2), retaining the log link and the assumption of constant coefficient of variation. These two models are equally effective over the range of temperatures observed but they exhibit rather different behaviour on extrapolation. Since the coefficient of variation for these data is so small, an equally effective analysis could be based on the logarithmic transformation of the observed durations.

From the point of view of gaining insight into the biochemistry, neither model (8.4) nor the polynomial model is very helpful. The biochemical mechanism of egg development appears to be rather complicated: a sequence or network of dependent biochemical reactions is likely to be involved. Furthermore, this experiment gives no information on the likely duration of development if the temperature were changed after, say, 5 hours. Powsner (p. 506) discusses the effects of such changes, which are quite complicated.

For a more recent review of the role of *Drosophila melanogaster* as an experimental organism, see the review article by Rubin (1988).

8.5 Bibliographic notes

There is an extensive literature on models for exponentially distributed observations. Such models are widely used for the distribution of lifetimes in industrial reliability experiments. See, for example, the books by Barlow and Proschan (1965, 1975) and Nelson (1982).

Similar models are used for survival analysis: details and further references are provided in Chapter 13.

The family of inverse linear models was introduced by Nelder (1966).

The gamma family, as parameterized here, has many properties in common with the Normal family. In particular, for a single sample, it is possible to construct exact similar regions for composite hypotheses specifying μ. Exact confidence intervals for μ can thereby be constructed, at least in principle. In practice the exact computations are excessively complicated for samples larger than about 3 or 4. For further details and references see Exercise 8.16.

8.6 Further results and exercises 8

8.1 Show that the standard deviation of $\log(Y)$ is approximately equal to the coefficient of variation of Y. Check numerically the adequacy of the approximation in the two cases

$$Y \sim G(\mu, \nu) \quad \text{and} \quad \log(Y) \sim N(\mu, \sigma^2)$$

for $\nu = 1, 2, \ldots$ and for various values of σ^2.

8.2 Show that the gamma distribution has cumulant generating function

$$K(t) = -\nu \log(1 - \mu t/\nu).$$

Hence deduce that for large ν, the standardized random variable $\nu^{1/2}(Y - \mu)/\mu$ is approximately distributed as $N(0, 1)$.

8.3 Assuming that Y has the gamma distribution, calculate the exact mean and variance of $\log(Y)$. Use the Tables in Abramowitz and Stegun (1970) to compare numerically these exact calculations with the approximate formulae in section 8.1.

8.4 Suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed with the gamma density $G(\mu, \nu)$. Show that $\bar{Y} = Y_{\cdot}/n$ is independent of $T = (Y_1/Y_{\cdot}, \ldots, Y_n/Y_{\cdot})$, and that the latter statistic has the symmetric Dirichlet distribution with index ν.

8.5 Show that for a simple random sample from the gamma distribution, the maximum-likelihood estimates of μ and ν are independent. Derive the conditional maximum-likelihood estimate of ν given $Y_{\cdot} = y_{\cdot}$. Compare this estimate with (8.2).

8.6 Fit the log-linear model

$$\text{PA} + \text{CG} + \text{VA}$$

to the insurance-claims data in Table 8.1. Use gamma errors and weight the averages according to the sample sizes. Examine the parameter estimates and state the conclusions to be drawn from your analysis as concisely as you can. Compare and contrast your conclusions with those presented in section 8.4.1.

8.7 For the data in Table 8.7 plot the log duration against the reciprocal of absolute temperature. Hence verify that the simple Arrhenius model does not fit these data.

8.8 Re-analyse the data in Table 8.7 using polynomials in place of (8.4). Try a number of link functions. Plot the residuals against temperature as a check on your model. Estimate the temperature at which the rate of development is a maximum.

8.9 Check whether the model described in section 8.4.4 might be improved by taking the response to be the time spent at the experimental temperature as opposed to the total duration of the embryonic period.

8.10 The data shown in Tables 8.9 and 8.10 were collected by Powsner (1935) in his study of the effect of temperature on the duration of the developmental stages of the fruit fly *Drosophila melanogaster*. In the light of the analyses suggested in section 8.4.4 for the embryonic period, examine carefully how the rates of development for the egg-larval, larval and pupal periods depend on temperature. What evidence is there of a maximum rate of development? Do the maximum developmental rates occur at the same temperature for each developmental stage? Check carefully whether there is any difference between the developmental rates for males and females, and if so, whether the difference is temperature-dependent.

8.11 By using the asymptotic expansion for $\psi(\nu)$, show that the maximum-likelihood estimate $\hat{\nu}$ in (8.1) is given approximately by

$$\frac{1}{\hat{\nu}} \simeq \frac{\bar{D}(6+\bar{D})}{6+2\bar{D}}$$

Table 8.9 *Mean duration of egg-larval and larval periods in the development of* Drosophila melanogaster

	Egg-larval period						*Larval period*		
Temp.	*Male*			*Female*			*Male*	*Female*	*Diff.*
°C	*Hours* ± s.e.	*No.*	σ	*Hours* ± s.e.	*No.*	σ	*Hours* ± s.e.	*Hours* ± s.e.	*F. − M.*
14.86	421.4 ± 1.68	97	16.3	423.8 ± 1.56	114	16.4	353.8 ± 1.6	356.2 ± 1.5	+2.4
15.24	394.6 ± 0.89	227	13.4	399.1 ± 0.82	227	12.2	330.6 ± 0.88	335.2 ± 0.79	+4.6
16.06	328.6 ± 0.91	148	11.0	342.2 ± 0.62	187	8.3	261.8 ± 0.90	285.5 ± 0.60	+13.7
18.04	250.1 ± 0.70	99	7.0	362.7 ± 0.76†	76	6.6	208.3 ± 0.69	222.0 ± 0.74	+13.7
18.05	241.3 ± 0.32	423	6.8	259.1 ± 0.32	407	6.6	199.5 ± 0.10	217.4 ± 0.10	+17.9
18.05	243.2 ± 0.53	178	7.1	262.4 ± 0.50	178	6.7	201.5 ± 0.52	220.7 ± 0.20	+19.2
18.21	242.2 ± 0.73	81	6.6	258.7 ± 0.60	84	5.5	201.3 ± 0.72	217.8 ± 0.58	+16.5
19.32	207.7 ± 0.36	122	3.9	222.2 ± 0.26	127	3.0	171.8 ± 0.11	186.4 ± 0.26	+14.6
19.97	188.0 ± 0.71	125	8.0	202.6 ± 0.82	138	9.7	154.5 ± 0.71	169.2 ± 0.82	+14.7
22.00	141.7 ± 0.18	128	2.0	149.9 ± 0.16	129	1.8	114.9 ± 0.17	123.2 ± 0.15	+8.3
22.00	141.6 ± 0.21	195	3.0	149.9 ± 0.17	183	2.3	114.8 ± 0.21	123.2 ± 0.16	+8.4
22.21	146.1 ± 0.31	139	3.6	153.1 ± 0.21	152	2.6	119.8 ± 0.30	126.9 ± 0.20	+7.1
22.99	130.5 ± 0.19	220	2.9	139.8 ± 0.22	216	3.2	105.9 ± 0.19	115.2 ± 0.22	+9.3
24.17	118.5 ± 0.29	140	3.4	120.7 ± 0.21	185	2.9	96.2 ± 0.29	98.4 ± 0.21	+2.2
24.93	113.3 ± 0.14	242	2.2	113.6 ± 0.16	229	2.4	92.4 ± 0.14	92.7 ± 0.16	+0.3
24.93	112.2 ± 0.13	341	2.4	113.4 ± 0.14	337	2.5	91.3 ± 0.13	92.5 ± 0.13	+1.2
25.14	113.7 ± 0.15	1004	4.6	114.2 ± 0.16	1039	5.2	93.1 ± 0.14	93.6 ± 0.16	+0.5
25.56	108.3 ± 0.14	357	2.6	108.3 ± 0.15	359	2.8	88.3 ± 0.14	88.3 ± 0.15	0.0
25.99	105.5 ± 0.14	344	2.6	105.8 ± 0.15	298	2.8	86.1 ± 0.14	86.4 ± 0.16	+0.3
26.89	100.1 ± 0.15	185	2.1	99.9 ± 0.18	203	2.6	81.5 ± 0.12	81.3 ± 0.18	−0.2
27.77	94.8 ± 0.14	128	1.5	94.3 ± 0.16	123	1.8	76.9 ± 0.13	76.4 ± 0.16	−0.5
27.77	94.4 ± 0.13	192	1.3	93.5 ± 0.12	207	1.7	76.5 ± 0.09	75.6 ± 0.12	−0.9
28.07	103.0 ± 0.34	74	2.9	102.9 ± 0.31	63	2.4	85.3 ± 0.32	85.3 ± 0.30	−0.1
28.99	98.4 ± 0.25	98	2.4	98.4 ± 0.28	96	2.7	81.1 ± 0.24	81.1 ± 0.27	0.0
29.47	99.2 ± 0.30	82	2.7	98.4 ± 0.32	88	3.0	82.1 ± 0.30	81.3 ± 0.32	−0.8
29.98	105.1 ± 0.25	226	3.7	105.1 ± 0.31	242	4.7	88.3 ± 0.24	88.3 ± 0.30	0.0
31.04	121.4 ± 0.56	157	7.1	118.5 ± 0.55	188	7.5	104.4 ± 0.57	101.4 ± 0.55	−3.0

†apparently misrecorded: should perhaps read 262.7 ± 0.76.

Source: Powsner (1935).

Table 8.10 *Mean duration of pupal period in the development of the fruit-fly* Drosophila melanogaster

Temp.	*Male*			*Female*			*Diff.*
°C	*Hours* ± s.e.	*No.*	σ	*Hours* ± s.e.	*No.*	σ	*M. − F.*
15.24	320.5 ± 0.45	228	6.72	309.5 ± 0.42	227	6.30	+11.0
16.17	266.7 ± 0.30	76	2.62	259.0 ± 0.24	186	3.16	+7.7
18.01	204.4 ± 0.28	97	2.78	195.3 ± 0.26	76	2.26	+9.1
18.05	204.4 ± 0.15	178	2.06	197.0 ± 0.14	174	1.81	+7.4
18.21	199.2 ± 0.22	83	2.01	192.2 ± 0.29	84	2.67	+7.0
19.32	170.6 ± 0.17	120	1.87	164.7 ± 0.10	126	1.12	+5.9
19.97	160.1 ± 0.21	125	2.40	152.2 ± 0.33	138	3.86	+7.9
22.00	126.5 ± 0.13	195	1.79	121.4 ± 0.14	182	1.91	+5.1
22.21	124.04 ± 0.089	138	1.05	120.70 ± 0.082	152	1.01	+3.34
22.99	115.62 ± 0.089	153	1.09	113.08 ± 0.074	215	1.10	+2.58
24.17	102.4 ± 0.14	140	1.64	98.65 ± 0.074	185	1.02	+3.75
24.57	100.51 ± 0.080	238	1.25	96.78 ± 0.065	229	1.00	+3.73
25.14	96.51 ± 0.044	967	1.37	93.55 ± 0.042	1018	1.35	+2.96
25.29	96.40 ± 0.096	99	0.96	92.23 ± 0.15	118	1.59	+4.17
25.99	92.24 ± 0.060	342	1.12	90.20 ± 0.058	298	1.01	+2.04
26.89	87.23 ± 0.075	185	1.02	82.56 ± 0.069	203	0.99	+4.67
27.77				80.0 ± 0.11	120	1.21	
28.07	83.2 ± 0.16	72	1.38	78.6 ± 0.13	64	1.07	+4.6
28.99	80.9 ± 0.11	85	0.99	76.2 ± 0.12	82	1.08	+4.7
29.47	80.6 ± 0.12	77	1.05	75.8 ± 0.11	70	0.92	+4.8
29.98	81.3 ± 0.11	233	1.68	77.1 ± 0.10	246	1.60	+4.2
30.24	82.0 ± 0.13	141	1.50	78.5 ± 0.12	161	1.50	+3.5
31.04	82.7 ± 0.17	73	1.4	79.3	22		+3.4

Source: Powsner (1935).

where $\tilde{D} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n$. Find the expected value of $\tilde{D}$ for an i.i.d. sample from the exponential distribution. Solve the above equation to find the approximate expected value of $\hat{\nu}$ when $\nu = 1$.

8.12 Show that the corresponding approximation for the bias-corrected estimate is

$$\frac{1}{\tilde{\nu}} \simeq \tilde{D}\frac{6(n-p)+n\tilde{D})}{6(n-p)+2n\tilde{D}}$$

where $\tilde{D} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p)$.

8.13 The data in Table 8.11 were obtained by Drs Streibig and Vleeshouwers in an experiment designed to study how the yields of various crops are affected by competition with weeds and by plant density. Taking the fresh weights as response, examine the relationship between the monoculture yields and seed density. (In

Table 8.11 *Yields of barley and the weed* Sinapis alba *grown in a competition experiment*†

	Seeds sown			*Plants harvested*		*Fresh weight*		*Dry weight*	
Pot	*Barley*	*Sinapis*	*Block*	*Barley*	*Sinapis*	*Barley*	*Sinapis*	*Barley*	*Sinapis*
1	3	0	1	3	0	33.7	0.0	2.07	0.00
2	5	0	1	5	0	120.5	0.0	10.57	0.00
3	7	0	1	7	0	187.3	0.0	20.87	0.00
4	10	0	1	10	0	110.1	0.0	6.59	0.00
5	15	0	1	15	0	122.7	0.0	8.08	0.00
6	23	0	1	23	0	214.9	0.0	16.70	0.00
7	34	0	1	33	0	198.6	0.0	21.22	0.00
8	51	0	1	48	0	263.6	0.0	26.57	0.00
9	77	0	1	60	0	254.1	0.0	23.71	0.00
10	115	0	1	70	0	230.4	0.0	20.46	0.00
11	0	5	1	0	5	0.0	254.0	0.00	34.85
12	3	5	1	3	5	14.8	167.6	1.49	29.49
13	7	5	1	6	5	38.1	240.5	2.26	19.75
14	15	5	1	15	5	93.1	132.6	11.08	23.09
15	34	5	1	34	5	120.8	166.9	12.85	25.83
16	77	5	1	50	5	214.5	53.2	24.94	8.76
17	0	7	1	0	7	0.0	228.3	0.00	38.98
18	0	10	1	0	10	0.0	209.8	0.00	28.14
19	3	10	1	3	11	15.2	220.1	1.63	35.43
20	7	10	1	7	7	37.6	203.0	2.80	29.05
21	15	10	1	15	10	93.3	130.5	6.29	17.36
22	34	10	1	31	9	98.6	178.5	7.81	23.30
23	77	10	1	62	10	203.9	81.5	19.51	12.45
24	0	15	1	0	16	0.0	214.4	0.00	36.02
25	0	23	1	0	22	0.0	269.3	0.00	47.24
26	3	23	1	3	23	7.5	272.2	1.06	49.14
27	7	23	1	6	23	18.8	220.1	1.83	35.85
28	15	23	1	14	23	64.7	175.8	9.35	30.05
29	34	23	1	28	25	84.3	240.3	9.75	36.46
30	77	23	1	53	26	125.8	135.5	14.29	19.87
31	0	34	1	0	33	0.0	267.4	0.00	38.23
32	0	51	1	0	53	0.0	244.6	0.00	31.75
33	3	51	1	1	58	3.3	332.0	0.34	35.68
34	7	51	1	7	54	21.5	264.9	2.11	37.95
35	15	51	1	11	53	26.4	221.5	1.89	25.78
36	34	51	1	23	50	32.8	230.1	3.97	39.97
37	77	51	1	61	52	76.9	184.0	7.60	24.42
38	0	77	1	0	81	0.0	291.9	0.00	45.56
39	0	115	1	0	115	0.0	300.3	0.00	43.94
40	3	115	1	3	108	1.3	284.9	0.13	29.39
41	7	115	1	6	109	5.8	243.7	0.55	33.44
42	15	115	1	14	111	12.1	287.9	0.95	35.68
43	34	115	1	26	107	26.4	233.0	2.07	21.53
44	77	115	1	57	115	95.9	189.2	10.14	24.02
45	0	173	1	0	158	0.0	326.4	0.00	35.24

(continued)

Table 8.11 *Continued*

	Seeds sown			*Plants harvested*		*Fresh weight*		*Dry weight*	
Pot	*Barley*	*Sinapis*	*Block*	*Barley*	*Sinapis*	*Barley*	*Sinapis*	*Barley*	*Sinapis*
46	3	0	2	3	0	73.1	0.0	5.32	0.00
47	5	0	2	5	0	152.7	0.0	13.59	0.00
48	7	0	2	7	0	125.4	0.0	9.97	0.00
49	10	0	2	10	0	208.9	0.0	21.40	0.00
50	15	0	2	15	0	171.5	0.0	11.07	0.00
51	23	0	2	19	0	98.7	0.0	6.66	0.00
52	34	0	2	27	0	191.8	0.0	14.25	0.00
53	51	0	2	41	0	238.7	0.0	39.37	0.00
54	77	0	2	49	0	197.2	0.0	21.44	0.00
55	115	0	2	72	0	256.4	0.0	30.92	0.00
56	0	5	2	0	5	0.0	227.3	0.00	32.61
57	3	5	2	3	5	28.9	246.3	1.66	34.18
58	7	5	2	8	5	42.3	230.0	3.62	33.63
59	15	5	2	15	5	82.9	156.1	10.41	27.06
60	34	5	2	28	5	116.7	125.9	10.46	19.99
61	77	5	2	57	5	187.7	55.8	23.10	9.01
62	0	7	2	0	7	0.0	231.5	0.00	34.20
63	0	10	2	0	10	0.0	258.8	0.00	44.47
64	3	10	2	3	10	25.8	245.1	2.88	38.13
65	7	10	2	7	11	41.6	185.8	5.32	33.31
66	15	10	2	14	11	67.7	174.3	11.10	32.49
67	34	10	2	33	11	86.0	177.9	9.16	29.20
68	77	10	2	65	10	162.3	75.0	23.18	12.52
69	0	15	2	0	15	0.0	237.6	0.00	40.32
70	0	23	2	0	29	0.0	225.9	0.00	37.46
71	3	23	2	3	29	9.8	274.0	0.76	46.68
72	7	23	2	6	27	27.7	221.4	1.96	34.40
73	15	23	2	13	23	30.2	246.0	4.19	45.07
74	34	23	2	24	25	110.0	147.9	11.34	22.75
75	77	23	2	56	24	85.7	185.4	10.39	30.38
76	0	34	2	0	34	0.0	281.0	0.00	43.76
77	0	51	2	0	51	0.0	318.9	0.00	40.25
78	3	51	2	5	54	6.8	309.5	0.12	45.80
79	7	51	2	7	51	12.1	254.2	0.26	29.29
80	15	51	2	13	51	28.5	226.8	2.32	31.28
81	34	51	2	25	52	55.4	195.3	5.73	29.61
82	77	51	2	49	50	109.7	179.0	8.61	20.14
83	0	77	2	0	76	0.0	304.7	0.00	38.94
84	0	115	2	0	101	0.0	313.5	0.00	43.36
85	3	115	2	3	99	10.6	249.2	0.97	36.54
86	7	115	2	10	109	7.4	255.9	0.01	31.80
87	15	115	2	16	97	17.9	170.9	1.05	24.96
88	34	115	2	22	105	38.1	270.9	3.65	38.52
89	77	115	2	47	97	52.5	266.6	6.40	38.91
90	0	173	2	0	148	0.0	279.2	0.00	39.35

(continued)

Table 8.11 *Continued*

	Seeds sown			*Plants harvested*		*Fresh weight*		*Dry weight*	
Pot	*Barley*	*Sinapis*	*Block*	*Barley*	*Sinapis*	*Barley*	*Sinapis*	*Barley*	*Sinapis*
91	3	0	3	3	0	42.9	0.0	3.14	0.00
92	5	0	3	5	0	165.9	0.0	14.69	0.00
93	7	0	3	7	0	81.4	0.0	5.45	0.00
94	10	0	3	9	0	223.3	0.0	23.12	0.00
95	15	0	3	17	0	116.3	0.0	8.28	0.00
96	23	0	3	20	0	193.7	0.0	19.48	0.00
97	34	0	3	29	0	237.1	0.0	38.11	0.00
98	51	0	3	42	0	264.0	0.0	25.53	0.00
99	77	0	3	47	0	241.0	0.0	19.72	0.00
100	115	0	3	73	0	269.0	0.0	41.02	0.00
101	0	5	3	0	5	0.0	184.2	0.00	36.18
102	3	5	3	3	5	22.9	142.2	1.86	23.39
103	7	5	3	7	5	58.0	166.6	8.37	31.13
104	15	5	3	16	10	77.4	181.6	7.97	29.64
105	34	5	3	32	8	114.8	141.6	14.14	24.57
106	77	5	3	53	8	124.7	86.4	16.37	15.46
107	0	7	3	0	11	0.0	235.9	0.00	35.44
108	0	10	3	0	14	0.0	200.8	0.00	33.60
109	3	10	3	3	15	12.6	197.7	1.50	37.66
110	7	10	3	9	14	46.7	231.1	5.61	39.01
111	15	10	3	16	14	44.5	198.2	4.05	29.21
112	34	10	3	27	15	73.4	122.1	8.33	20.35
113	77	10	3	53	12	132.4	121.9	20.59	23.64
114	0	15	3	0	15	0.0	251.8	0.00	32.16
115	0	23	3	0	26	0.0	229.3	0.00	28.46
116	3	23	3	3	24	9.0	244.8	0.30	33.31
117	7	23	3	7	28	27.1	203.8	3.39	35.58
118	15	23	3	14	24	37.0	218.2	2.31	30.36
119	34	23	3	22	35	71.2	152.3	7.35	21.86
120	77	23	3	57	22	91.9	158.1	13.45	27.69
121	0	34	3	0	42	0.0	250.9	0.00	41.58
122	0	51	3	0	53	0.0	275.6	0.00	46.00
123	3	51	3	3	52	5.1	298.8	0.41	44.17
124	7	51	3	7	51	15.4	257.2	1.45	38.54
125	15	51	3	13	53	32.1	232.0	2.50	31.47
126	34	51	3	27	51	37.1	191.0	2.35	19.96
127	77	51	3	60	51	107.6	179.8	10.04	22.79
128	0	77	3	0	79	0.0	286.6	0.00	44.71
129	0	115	3	0	111	0.0	289.6	0.00	38.34
130	3	115	3	3	103	3.5	236.9	0.06	27.80
131	7	115	3	8	113	13.8	239.0	1.25	34.50
132	15	115	3	14	110	11.6	289.7	0.59	28.42
133	34	115	3	26	90	38.7	246.6	2.80	27.79
134	77	115	3	56	82	78.0	194.6	8.67	26.64
135	0	173	3	0	152	0.0	252.7	0.00	24.41

[†] Data courtesy of Drs J.C. Streibig and L. Vleeshouwers, Dept. of Crop Science, Royal Veterinary and Agricultural University, Copenhagen.

a few cases there was some doubt concerning the temperature of the drying process, so that the fresh weights may be more reliable than the dry weights.) Show that, for both barley and *Sinapis*, the relation between monoculture yield and seed density is approximately inverse quadratic. The monoculture observations are that subset of Table 8.11 in which only one variety was planted.

8.14 For those plots in which both varieties were sown, examine how the barley proportion of the total yield depends on the barley proportion of the seeds sown, the seed density and the experimental block. Take the log ratio of fresh weights, $\log(Y_{\text{B}}/Y_{\text{S}})$, as the response and consider models of the form

$$\log(Y_{\text{B}}/Y_{\text{S}}) = \alpha_{\text{BS}} + \beta\log(N_{\text{B}}/N_{\text{S}}) + \gamma x + \text{block effect},$$

where $N_{\text{B}}, N_{\text{S}}$ are the numbers of seeds sown, and x is a measure of seed density, say $x = \log(N_{\text{B}} + N_{\text{S}})$.

What would be the interpretation of the following parameter values?

1. $\beta = 1, \gamma = 0$;
2. $\beta < 1, \gamma = 0$;
3. $\beta = 1, \gamma > 0$;
4. $\beta < 1, \gamma > 0$.

For further information concerning competition experiments, see Williams (1962), Breese and Hill (1973), Mead and Curnow (1983) or Skovgaard (1986).

8.15 Suppose that $Y_i \sim G(\mu_i, \nu)$ independently for each i, with μ_i satisfying the log-linear model

$$\log(\mu_i) = \alpha + \mathbf{x}_i^T\boldsymbol{\beta}$$

and ν an unknown constant. Show that the transformed responses satisfy

$$E\big(\log(Y_i)\big) = \alpha^* + \mathbf{x}_i^T\boldsymbol{\beta},$$
$$\operatorname{var}(\log Y_i) = \psi'(\nu).$$

where $\alpha^* = \alpha + \psi(\nu) - \log(\nu)$ and $\psi(\nu) = \Gamma'(\nu)/\Gamma(\nu)$.

Let $\tilde{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$ obtained by fitting a linear regression model to the transformed data. Show that $\tilde{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ and that the asymptotic efficiency of $\tilde{\boldsymbol{\beta}}$ relative to $\hat{\boldsymbol{\beta}}$ is $1/\{\nu\psi'(\nu)\}$. [Bartlett and Kendall 1946; Cox and Hinkley 1966].

8.16 Suppose that $Y_i, \ldots, Y_n$ are independent and identically distributed with the gamma distribution $G(\mu, \nu)$, both parameters being taken as unknown. Let $\bar{Y}$ be the arithmetic mean of the observations, and $\dot{Y}$ the geometric mean. Show that under the composite hull hypothesis $H_0\colon \mu = \mu_0$, $S_0 = \log(\dot{Y}/\mu_0) - \bar{Y}/\mu_0$ is a complete sufficient statistic for ν. Show, again under H_0, that the conditional joint distribution of $Z_i = \log(Y_i/\mu_0)$ given S_0 is uniform over the surface

$$\sum \bigl(z_i - \exp(z_i)\bigr) = nS_0.$$

Discuss briefly how you might use this result (i) to construct an exact test of H_0, and (ii) to construct an exact confidence interval for μ.