

MATH 523
GENERALIZED LINEAR MODELS

WINTER 2013

MATH 523: GENERALIZED LINEAR MODELS

Instructor : David A. Stephens (Burnside 1225)
Email : d.stephens@math.mcgill.ca
Website: www.math.mcgill.ca/~dstephens/523/
Office Hours : Tuesday 10:30-12:00
Thursday 10:30-12:00
Other appointments available if necessary
Textbooks : *Categorical Data Analysis* (3rd Ed.) by A. Agresti,
Generalized Linear Models (2nd Ed) by P. McCullagh & J.A. Nelder,
Evaluation : Assignments (15 %), Mid Term (25 %), Final (60 %)
See p. ii for full details

TARGET SYLLABUS

1 Review of the linear model

- 1.1 least-squares estimation
- 1.2 distribution of estimators
- 1.3 hypothesis testing
- 1.4 aliasing and multicollinearity
- 1.5 interactions
- 1.6 response transformations
- 1.7 residuals

2 Generalized linear models (GLMs)

- 2.1 exponential families
- 2.2 link functions
- 2.3 maximum likelihood
- 2.4 iteratively weighted least-squares
- 2.5 quadratic algorithms for maximum likelihood
- 2.6 asymptotic distribution of maximum likelihood estimators; likelihood ratio tests; deviance
- 2.7 residuals for GLMs

3 Models for count data

- 3.1 Poisson regression and log-linear models
- 3.2 Implementation in R
- 3.3 goodness-of-fit tests
- 3.4 contingency table analysis
- 3.5 structured log-linear models: independence and symmetry
- 3.6 iterative proportional fitting
- 3.7 overdispersion and underdispersion

4 Models for binomial and multinomial data

- 4.1 binomial and logistic regression
- 4.2 logistic regression as a special case of log-linear modelling
- 4.3 2×2 tables and case-control designs
- 4.4 overdispersion for binomial data
- 4.5 multinomial responses
- 4.6 conditional logistic regression
- 4.7 matched pairs

5 Special topics

- 5.1 gamma regression
- 5.2 quasi-likelihood and generalized estimating equations
- 5.3 estimation of the link function

EVALUATION: DETAILS

Please note that the method of evaluation for this class will be **on the following basis only**:

Coursework Assignments	Four Assignments Beginning Tuesday 22nd January 2013
------------------------	---

Mid-Term	1.5 hours Thursday 21st February 2013 In class Closed book
----------	---

Final	3 hours Closed book
-------	------------------------

Final mark for course[‡]: the larger of

15 % Coursework + 25 % Mid-Term + 60 % Final

and

15 % Coursework + 85 % Final

NOTES:

- ‡ There will no opportunity for a make-up Mid-Term if this examination is missed, and no make-up work in place of any aspect of the course assessment. Assignments handed in late will be subject to penalties of up to 100 %

ACADEMIC INTEGRITY

McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see

<http://www.mcgill.ca/integrity/>

for more information).

- In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.
- In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.

David A. Stephens.
17th December 2012.

1 REVIEW OF THE LINEAR MODEL

1.1 KEY RESULTS FOR THE LINEAR MODEL

The standard linear model formulation suitable for representing the variation in data $y = (y_1, \dots, y_n)^\top$ sets

$$Y = \mathbf{X}\beta + \varepsilon$$

where

- Y is the $n \times 1$ vector of response random variables;
- $\mathbf{X}$ is the $n \times p$ design matrix (regarded as a constant, non-random quantity) with
 - rows $x_i, i = 1, \dots, n$, where x_i is $p \times 1$ covariate vector for datum i ,
 - columns $x_j, j = 1, \dots, p$, where x_j is $n \times 1$ vector of values for covariate j ;
- β is a $p \times 1$ vector of parameters;
- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ is an $n \times 1$ vector of random residual error variables with

$$\mathbb{E}[\varepsilon] = \mathbf{0}_n \quad \text{Var}[\varepsilon] = \Sigma < \infty$$

and most commonly $\Sigma = \sigma^2 \mathbf{I}_n$, so that the elements of ε are identically distributed and uncorrelated. Often, the assumption

$$\varepsilon \sim \mathcal{N}_n(0, \Sigma)$$

is made.

Estimation for β proceeds by least-squares (or maximum likelihood estimation if distributional assumptions are made about ε). Under the least-squares criterion (or ML if the Gaussian distributional assumption is made), we have that the estimate satisfies

$$\hat{\beta} = \arg \min_{\beta} (y - \mathbf{X}\beta)^\top \Sigma^{-1} (y - \mathbf{X}\beta) = \arg \min_{\beta} S(\beta)$$

say. For estimation, the *normal* equations (or *score* equations) are obtained by differentiating $S(\beta)$ with respect to β , and equating to zero. In vector form, we have for the estimator

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}^\top \Sigma^{-1} Y + 2\mathbf{X}^\top \Sigma^{-1} \mathbf{X} \beta = \mathbf{0}_p \quad (p \times 1)$$

so that

$$\mathbf{X}^\top \Sigma^{-1} \mathbf{X} \hat{\beta} = \mathbf{X}^\top \Sigma^{-1} Y.$$

If the $p \times p$ matrix $\mathbf{X}^\top \Sigma^{-1} \mathbf{X}$ is non-singular, we have that

$$\hat{\beta} = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} Y$$

If $\Sigma = \sigma^2 \mathbf{I}_n$, we have that

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y.$$

1.2 PROPERTIES OF THE ESTIMATORS

We have for the properties of the estimator $\hat{\beta}$ that

$$\mathbb{E}[\hat{\beta}] = \beta \quad \text{Var}[\hat{\beta}] = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1}$$

If $\Sigma = \sigma^2 \mathbf{I}_n$, then

$$\text{Var}[\hat{\beta}] = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$$

where σ^2 is commonly estimated by

$$\hat{\sigma}^2 = \frac{1}{n-p} (y - \mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta}).$$

If $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathbf{I}_n)$, then we have for the sampling distribution of $\hat{\beta}$

$$\hat{\beta} \sim \mathcal{N}_n(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$$

1.3 HYPOTHESIS TESTING

Under the Gaussianity assumption, or approximately for large n , under the assumption of uncorrelated residual errors, we have the following methods of hypothesis testing:

(a) **Tests for individual coefficients:** To test

$$\begin{aligned} H_0 &: \beta_j = b_j \\ H_1 &: \beta_j \neq b_j \end{aligned}$$

we have that

$$\frac{\hat{\beta}_j - b_j}{\sqrt{v_{jj}}} \sim \text{Student}(n-p)$$

where v_{jj} is the j th diagonal element of $\text{Var}[\hat{\beta}]$.

(b) **Tests for linear restrictions:** To test

$$\begin{aligned} H_0 &: A\beta = c \\ H_1 &: A\beta \neq c \end{aligned}$$

where H_0 places q linear constraints on the parameters of the model, we may use the following F -test. If $\hat{\beta}_0$ is the parameter estimate under H_0 , and $\hat{\beta}_1$ is the parameter estimate under $H_0 \cup H_1$, and

$$\begin{aligned} S(\hat{\beta}_0) &= (Y - \mathbf{X}\hat{\beta}_0)^\top (Y - \mathbf{X}\hat{\beta}_0) \\ S(\hat{\beta}_1) &= (Y - \mathbf{X}\hat{\beta}_1)^\top (Y - \mathbf{X}\hat{\beta}_1) \end{aligned}$$

then, under H_0 , the test statistic (random variable)

$$F = \frac{[S(\hat{\beta}_0) - S(\hat{\beta}_1)]/q}{S(\hat{\beta}_1)/(n-p)}$$

has a Fisher-F distribution, $F \sim \text{Fisher}(q, n-p)$.

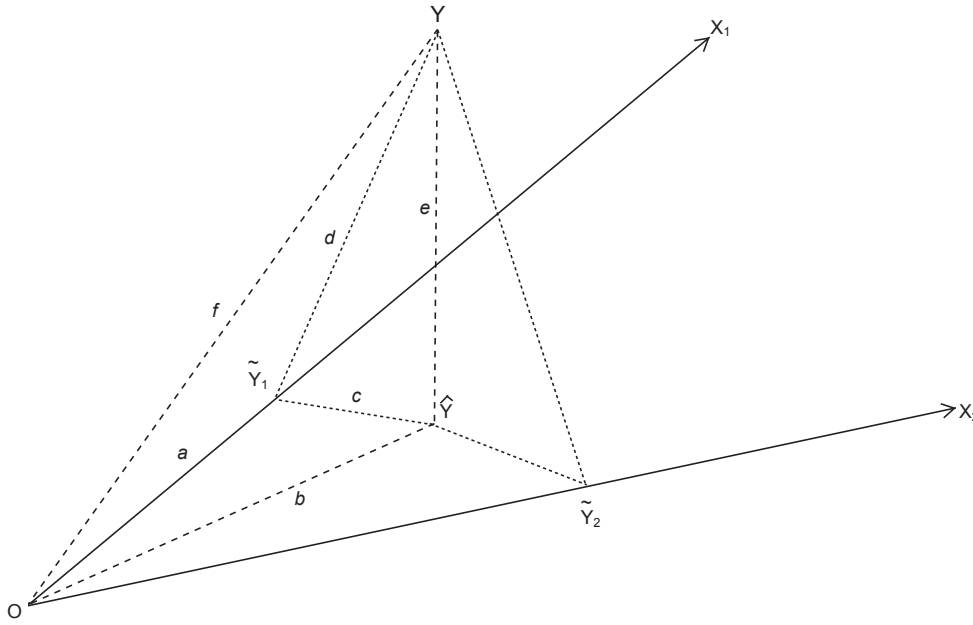
Note, by properties of the Fisher-F distribution, that as $n \rightarrow \infty$,

$$F \xrightarrow{d} Q \sim \frac{1}{q} \chi_q^2$$

1.4 THE GEOMETRY OF MULTICOLLINEARITY AND SUMS OF SQUARES FOR THE LINEAR MODEL.

Consider a two covariate linear model with (collections of) covariates X_1 and X_2 . Let

- $\tilde{Y}_1 = X_1\tilde{\beta}_1$ be the least-squares fit using only X_1 .
- $\tilde{Y}_2 = X_2\tilde{\beta}_2$ be the least-squares fit using only X_2 .
- $\hat{Y} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2$ be the least-squares fit using both X_1 and X_2 .



so that, for example

Vector	$O\tilde{Y}_1$	$O\hat{Y}$	$\tilde{Y}_1\hat{Y}$	$\tilde{Y}_1Y$	$\hat{Y}Y$	OY
Length	a	b	c	d	e	f

By previous orthogonality results,

- (i) $\tilde{Y}_1$ is the projection of Y onto the X_1 subspace, hence the angle $O\tilde{Y}_1Y$ is a right angle, so

$$\|O\tilde{Y}_1\|^2 + \|\tilde{Y}_1Y\|^2 = \|OY\|^2$$

$$\text{or } a^2 + d^2 = f^2.$$

- (ii) $\hat{Y}$ is the projection of Y onto the (X_1, X_2) subspace, hence the angle $O\hat{Y}Y$ is a right angle, so

$$\|O\hat{Y}\|^2 + \|\hat{Y}Y\|^2 = \|OY\|^2$$

$$\text{or } b^2 + e^2 = f^2.$$

However, $\tilde{Y}_1\hat{Y}Y$ is also a right angle, so $c^2 + e^2 = d^2$. Hence we must have also that $a^2 + c^2 = b^2$, and can deduce that $O\tilde{Y}_1\hat{Y}$ is also a right angle. Thus $\tilde{Y}_1$ is the projection of $\hat{Y}$ onto the X_1 space. By similar arguments, $\tilde{Y}_2$ is the projection of $\hat{Y}$ onto the X_2 space.

Two sequential model fitting strategies are possible:

(I) Fit X_1 , then fit X_2 to the residual after fitting X_1 . This yields the sum of squares decomposition

$$\begin{aligned}\|OY\|^2 &= \|O\tilde{Y}_1\|^2 + \|\tilde{Y}_1\hat{Y}\|^2 + \|\hat{Y}Y\|^2 \\ Y^\top Y &= \tilde{\beta}_1^\top X_1^\top X_1 \tilde{\beta}_1 + (X_1\hat{\beta}_1 + X_2\hat{\beta}_2 - X_1\tilde{\beta}_1)^\top (X_1\hat{\beta}_1 + X_2\hat{\beta}_2 - X_1\tilde{\beta}_1) \\ &\quad + (Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)^\top (Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)\end{aligned}$$

(II) Fit X_2 , then fit X_1 to the residual after fitting X_2 . This yields the sum of squares decomposition

$$\begin{aligned}\|OY\|^2 &= \|O\tilde{Y}_2\|^2 + \|\tilde{Y}_2\hat{Y}\|^2 + \|\hat{Y}Y\|^2 \\ Y^\top Y &= \tilde{\beta}_2^\top X_2^\top X_2 \tilde{\beta}_2 + (X_1\hat{\beta}_1 + X_2\hat{\beta}_2 - X_2\tilde{\beta}_2)^\top (X_1\hat{\beta}_1 + X_2\hat{\beta}_2 - X_2\tilde{\beta}_2) \\ &\quad + (Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)^\top (Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)\end{aligned}$$

The terms $\|OY\|^2$ and $\|\hat{Y}Y\|^2$ are common to both expressions, and independent of the strategy. Thus, in the usual terminology, we have two equivalent expressions for the *fitted sum of squares*

$$\|O\tilde{Y}_1\|^2 + \|\tilde{Y}_1\hat{Y}\|^2 = \|O\tilde{Y}_2\|^2 + \|\tilde{Y}_2\hat{Y}\|^2$$

Using the standard formulae, we have that if $\mathbf{X} = [X_1^\top, X_2^\top]^\top$, then

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} X_1^\top X_1 & X_1^\top X_2 \\ X_2^\top X_1 & X_2^\top X_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad \therefore \quad (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$$

say, where

$$A = (S_{11} - S_{12}S_{22}^{-1}S_{21})^{-1} \quad B = -S_{11}S_{12}(S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1} \quad C = (S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1}$$

Therefore, for the model with both covariates, we have

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y = \begin{bmatrix} AX_1^\top Y + BX_2^\top Y \\ B^\top X_1^\top Y + CX_2^\top Y \end{bmatrix}$$

so that

$$\begin{aligned}\hat{\beta}_1 &= (X_1^\top X_1 - (X_1^\top X_2)(X_2^\top X_2)^{-1}(X_2^\top X_1))^{-1} X_1^\top Y \\ &\quad - \{(X_1^\top X_1)^{-1}(X_1^\top X_2)(X_2^\top X_2 - (X_2^\top X_1)(X_1^\top X_1)^{-1}(X_1^\top X_2))\} X_2^\top Y\end{aligned}$$

whereas for the model with just X_1

$$\tilde{\beta}_1 = (X_1^\top X_1)^{-1} X_1^\top Y.$$

Therefore if $X_1^\top X_2 \neq 0$, $\hat{\beta}_1 \neq \tilde{\beta}_1$. If $X_1^\top X_2 = 0$, so that X_1 and X_2 are orthogonal, then in the figure, the angle

$$\tilde{Y}_1 O \tilde{Y}_2$$

is also a right angle.

1.5 INTERACTIONS

TWO-WAY INTERACTIONS:

An interaction between predictors X_1 and X_2 is present in a model if the effect of changing X_1 on the response Y is **different** for different values of X_2 .

- (1) **FACTORS:** For two **factor** predictors, an interaction exists if

$$\begin{aligned} \mathbb{E}[Y|X_1 = x_{11}, X_2 = x_{21}] - \mathbb{E}[Y|X_1 = x_{12}, X_2 = x_{21}] \\ \neq \mathbb{E}[Y|X_1 = x_{11}, X_2 = x_{22}] - \mathbb{E}[Y|X_1 = x_{12}, X_2 = x_{22}] \end{aligned}$$

for at least one choice of factor levels $(x_{11}, x_{12}, x_{21}, x_{22})$. For example, if X_1 and X_2 have two levels, and the means for the responses for different factor level combinations are

		X_2	
		1	2
X_1	1	0	-3
	2	2	4

then

$$\begin{aligned} \mathbb{E}[Y|X_1 = 1, X_2 = 1] - \mathbb{E}[Y|X_1 = 2, X_2 = 1] &= 0 - 2 = -2 \\ \mathbb{E}[Y|X_1 = 1, X_2 = 2] - \mathbb{E}[Y|X_1 = 2, X_2 = 2] &= (-3) - 4 = -7 \end{aligned}$$

- (2) **COVARIATES:** If

$$\mathbb{E}[Y|X_1 = x_1, X_2 = x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

then

$$\mathbb{E}[Y|X_1 = x_{11}, X_2 = x_2] - \mathbb{E}[Y|X_1 = x_{12}, X_2 = x_2] = \beta_1(x_{11} - x_{12}) + \beta_{12}(x_{11} - x_{12})x_2$$

which is a function of x_2 .

- (3) **FACTORS and COVARIATES:** Suppose X_1 is a factor taking L levels, and X_2 is a covariate. Then $X_{1l} = l$ say defines a subset of the subjects; interaction models allow the modelling of the responses as a function of the covariate to be different for different values of l . For example, the model written

$$1 + X_1 + X_2 + X_1.X_2$$

is a *main effects plus interaction* model for which

$$\mathbb{E}[Y|X_1 = l, X_2 = x_2] = (\beta_0 + \beta_{0l}) + (\beta_1 + \beta_{1l})x_2$$

for $l = 2, \dots, L$.

HIGHER-ORDER INTERACTIONS

Higher-order interactions represent the changes in lower-order interaction effects at different levels of predictors. For example, for three predictors X_1, X_2, X_3 , it is possible to include three **two-way** interactions

$$X_1.X_2 \quad X_1.X_3 \quad X_2.X_3$$

but also a **three-way** interaction, $X_1.X_2.X_3$, which represents the changes in the values in each of the two-way interactions at different levels/values of the third variable.

INTERACTIONS WITHOUT MAIN EFFECTS

A commonly adopted convention is to only consider models for which, if a M -way interaction is included, then all m -way interactions between the M variables involved are also included, for $m = 2, 3, \dots, M - 1$. That is, if $M = 4$ and we wish to include a 4-way interaction, we should also include all the

$$\begin{aligned}\binom{4}{3} &= 4 \quad \text{three-way interactions} \\ \binom{4}{2} &= 6 \quad \text{two-way interactions} \\ \binom{4}{1} &= 4 \quad \text{main effects}\end{aligned}$$

Such a model is termed a *full factorial* model.

NOTE: Unless there is a specific need or reason to do so, **we do not omit any** lower-order interactions or main effects. That is, for a two-predictor model, we might consider the full factorial model

$$1 + X_1 + X_2 + X_1.X_2$$

but **not** the model

$$1 + X_1 + X_1.X_2$$

which omits the X_2 main effect. The implication of the second model is that X_2 is influential in explaining the response, but only for **specific** values of X_1 . For example, for two factor predictors with J and K levels, the second model might imply

$$\mathbb{E}[Y|X_1 = j, X_2 = k] = \begin{cases} \beta_0 & j = 1, k = 1, \dots, K \\ \beta_0 + \beta_{1j} & j = 2, \dots, J, k = 1 \\ \beta_0 + \beta_{1j} + \gamma_{jk} & j = 2, \dots, J, k = 1, \dots, K \end{cases}$$

so that at level $X_1 = 1$, changing X_2 has **no influence** on the expected response, whereas for $j \geq 2$, changing X_2 **does** have an influence.

NOTE: A single m -way interaction term may correspond to **many** individual parameters. The general rules for counting parameters for main effects and interactions in a factorial model are established as follows:

Term	Type	No. of Levels	No. of Extra Parameters
Intercept			1
Main effect	Factor	X_1	$M - 1$
	Covariate		1
2-way interaction	Two Factors	X_1, X_2	$(M_1 - 1)(M_2 - 1)$
	Two covariates		1
	Factor and covariate		$(M_1 - 1).1$
3-way interaction	Three Factors	X_1, X_2, X_3	$(M_1 - 1)(M_2 - 1)(M_3 - 1)$
	Three covariates		1

and so on.

1.6 RESIDUALS

Residual checking is a key component of assessment of model fit for the linear model. The vector of fitted values is $\hat{Y} = \mathbf{X}\hat{\beta}$, and the $(n \times 1)$ vector of residuals is

$$R = Y - \hat{Y} = Y - \mathbf{X}\hat{\beta} = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)Y = (\mathbf{I}_n - \mathbf{H})Y$$

with observed values

$$r = y - \hat{y} = (\mathbf{I}_n - \mathbf{H})y.$$

Under the assumption of a correct model specification, we have that

$$\begin{aligned} \hat{y}^\top r &= (\mathbf{X}\hat{\beta})^\top (y - \mathbf{X}\hat{\beta}) = \hat{\beta}^\top \mathbf{X}^\top y - \hat{\beta}^\top (\mathbf{X}^\top \mathbf{X})\hat{\beta} \\ &= y^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top y - y^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X})(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top y \\ &= 0 \end{aligned}$$

by the previous orthogonality results. By similar arguments, we have that

$$\mathbf{X}^\top r = \mathbf{X}^\top (y - \mathbf{X}\hat{\beta}) = \mathbf{0}_p$$

that is, the observed residuals are

- orthogonal to fitted values $\hat{y}$
- orthogonal to the columns of $\mathbf{X}$.

so therefore plots of residuals versus fitted values or the columns of $\mathbf{X}$ should show no pattern, and be centered at zero.

Note, however, that

$$y^\top r = y^\top (y - \mathbf{X}\hat{\beta}) = y^\top y - y^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top y = y^\top (\mathbf{I}_n - \mathbf{H})y \neq 0$$

in general, so the residuals are **not** orthogonal to the observed responses.

Note also that by similar calculations for the random variables

$$\mathbb{E}[\hat{Y}^\top R] = 0 \quad \mathbb{E}[\mathbf{X}^\top R] = \mathbf{0}_p$$

and that

$$\text{Var}[R] = \text{Var}[(\mathbf{I}_n - \mathbf{H})Y] = (\mathbf{I}_n - \mathbf{H})\Sigma(\mathbf{I}_n - \mathbf{H})^\top = \mathbf{V}_R$$

say. This suggests that the residuals can be *standardized* by dividing them by their (estimated) standard errors; the standard error for R_i is the square root of the (i, i) element of the matrix $\mathbf{V}_R$

$$r_i^{(S)} = \frac{y_i - (\mathbf{X}\hat{\beta})_i}{\sqrt{V_{Rii}}}.$$

If $\Sigma = \sigma^2 \mathbf{I}_n$, then

$$\mathbf{V}_R = \sigma^2 (\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})^\top = \sigma^2 (\mathbf{I}_n - \mathbf{H} - \mathbf{H} + \mathbf{H}\mathbf{H}^\top) = \sigma^2 (\mathbf{I}_n - \mathbf{H})$$

as $\mathbf{H}\mathbf{H}^\top = \mathbf{H}$ so that the i^{th} standardized residual is

$$r_i^{(S)} = \frac{y_i - (\mathbf{X}\hat{\beta})_i}{\hat{\sigma}\sqrt{1 - H_{ii}}}.$$

2 GENERALIZED LINEAR MODELS

2.1 EXPONENTIAL FAMILY EXAMPLES

The mass/density function f_Y falls into the Exponential-Dispersion family with scalar parameters (θ, ϕ) if f_Y can be written

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}$$

for specific functions a , b and c , and the support of the function does not depend on (θ, ϕ) .

- θ is the **canonical** parameter
- $b(\theta)$ is the **cumulant function** parameter
- ϕ is the **dispersion** parameter

Note that the function $a(\cdot)$ is often the identity, that is

$$a(\phi) = \phi.$$

If

$$l(\theta, \phi; y) = \log f_Y(y; \theta, \phi)$$

it follows by direct calculations that

$$\mathbb{E}_{f_Y}[\dot{l}_\theta(\theta, \phi; Y)] = 0$$

and

$$\mathbb{E}_{f_Y}[\ddot{l}_{\theta\theta}(\theta, \phi; Y)] + \mathbb{E}_{f_Y}\left[\left\{\dot{l}_\theta(\theta, \phi; Y)\right\}^2\right] = 0$$

where

$$\dot{l}_\theta(\theta, \phi; y) = \frac{\partial l(\theta, \phi; y)}{\partial \theta} \quad \ddot{l}_{\theta\theta}(\theta, \phi; y) = \frac{\partial^2 l(\theta, \phi; y)}{\partial \theta^2}$$

Hence in the Exponential Family, we have

$$\mathbb{E}[Y] = \dot{b}(\theta) = \mu \quad \text{Var}[Y] = a(\phi)\ddot{b}(\theta) = a(\phi)V(\mu)$$

say.

- $\mu = \dot{b}(\theta)$ is the **mean** function
- $V(\mu) = \ddot{b}(\theta)a(\phi)$ is the **variance** function.

POISSON For $y = 0, 1, 2, \dots$

$$f_Y(y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\{y \log \lambda - \lambda - \log y!\}$$

- Canonical parameter: $\theta = \log \lambda$
- Cumulant function: $b(\theta) = e^\theta$
- Dispersion parameter: $a(\phi) = 1$
- Mean function: $\mathbb{E}[Y] = \dot{b}(\theta) = e^\theta$
- Variance function: $\text{Var}[Y] = \ddot{b}(\theta)a(\phi) = e^\theta = \mu$, so $V(\mu) = \mu$.

BERNOULLI For $y = 0, 1$

$$f_Y(y; \pi) = \pi^y(1 - \pi)^{1-y} = \exp \left\{ y \log \left(\frac{\pi}{1 - \pi} \right) + \log(1 - \pi) \right\}$$

- Canonical parameter: $\theta = \log \left(\frac{\pi}{1 - \pi} \right)$
- Cumulant function: $b(\theta) = \log(1 + e^\theta)$
- Dispersion parameter: $a(\phi) = 1$
- Mean function: $\mathbb{E}[Y] = \dot{b}(\theta) = \frac{e^\theta}{1 + e^\theta} = \pi$
- Variance function: $\text{Var}[Y] = \ddot{b}(\theta)a(\phi) = \frac{e^\theta}{(1 + e^\theta)^2} = \mu(1 - \mu)$, so $V(\mu) = \mu(1 - \mu)$.

BINOMIAL For $z = 0, 1, \dots, m$

$$f_Z(z; \pi) = \binom{m}{z} \pi^z (1 - \pi)^{m-z} = \exp \left\{ z \log \left(\frac{\pi}{1 - \pi} \right) + m \log(1 - \pi) + \log \binom{m}{z} \right\}$$

Consider the data transformation $y = z/m$. Then for $y = 0, 1/m, 2/m, \dots, 1$ we have

$$f_Y(y; \pi) = \exp \left\{ \frac{y \log \left(\frac{\pi}{1 - \pi} \right) + \log(1 - \pi)}{1/m} + \log \binom{m}{my} \right\}$$

- Canonical parameter: $\theta = \log \left(\frac{\pi}{1 - \pi} \right)$
- Cumulant function: $b(\theta) = \log(1 + e^\theta)$
- Dispersion parameter: $a(\phi) = \phi = 1/m$
- Mean function: $\mathbb{E}[Y] = \dot{b}(\theta) = \frac{e^\theta}{1 + e^\theta} = \pi$
- Variance function: $\text{Var}[Y] = \ddot{b}(\theta)a(\phi) = \frac{e^\theta}{m(1 + e^\theta)^2} = \mu(1 - \mu)/m$, so $V(\mu) = \mu(1 - \mu)$.

GAMMA For $y > 0$

$$f_Y(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} = \exp \{ (\alpha - 1) \log y - \beta y + \alpha \log \beta - \log \Gamma(\alpha) \}$$

Consider the reparameterization

$$\nu = \alpha \quad \mu = \alpha/\beta$$

so that

$$f_Y(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu y}{\mu} \right)^\nu \exp \left\{ -\frac{\nu y}{\mu} \right\} \frac{1}{y}$$

and we have

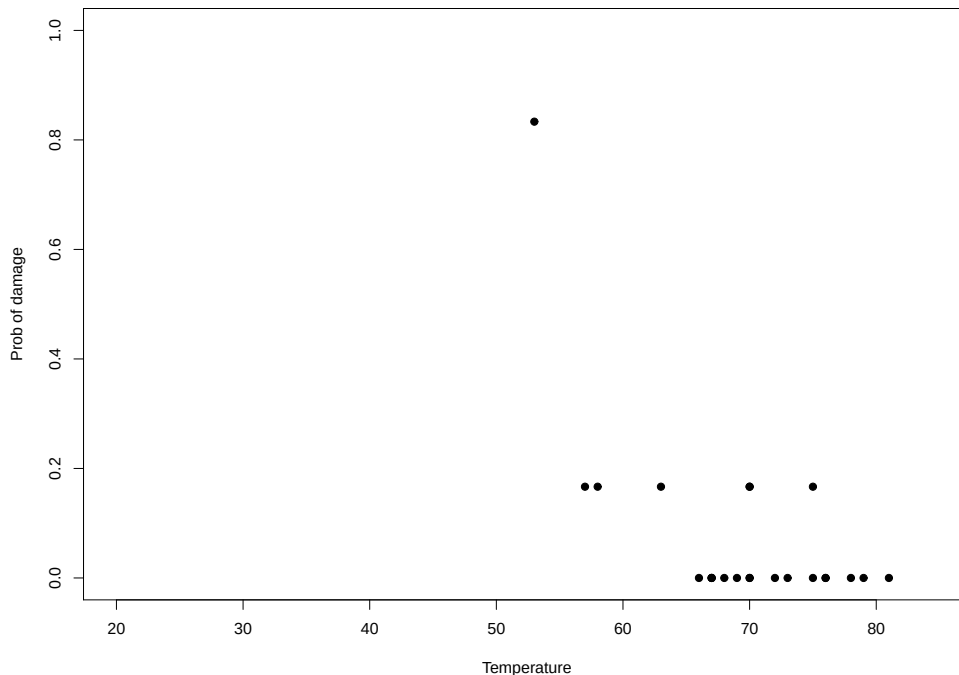
- Canonical parameter: $\theta = -1/\mu$.
- Cumulant function: $b(\theta) = -\log(-\theta)$
- Dispersion parameter: $a(\phi) = \phi = 1/\nu$.
- Mean function: $\mathbb{E}[Y] = \dot{b}(\theta) = -1/\theta$
- Variance function: $\text{Var}[Y] = \ddot{b}(\theta)a(\phi) = (1/(1/\mu^2))(1/\nu)$, so $V(\mu) = \mu^2$.

EXAMPLE: BINOMIAL DATA

The following data are counts of the number of failed “O-rings” (out of six) in twenty three launches of the space shuttle Challenger at different ground temperatures ranging from 53F to 81F. We seek to assess the probability of failure of all six rings at different temperatures to assess the safety.

Temperature (F)	53	57	58	63	66	67	67	67	68	69	70	70	70	70	72	73	75	75	76	76	78	79	81
No. damaged (out of 6)	5	1	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0

The following plot depicts the proportion of failures at the different temperatures.



Consider three different models: if Y is the proportion of failures, we fit

- A linear model:

$$\mathbb{E}[Y_i] = \beta_0 + \beta_1 T_i$$

- A GLM with **logistic** link: if $\text{expit}(t) = e^t / (1 + e^t)$

$$\mathbb{E}[Y_i] = \text{expit}(\beta_0 + \beta_1 T_i)$$

- A GLM with **probit** link: if $\Phi(t)$ is the standard normal CDF

$$\mathbb{E}[Y_i] = \Phi(\beta_0 + \beta_1 T_i)$$

The parameters can be estimated in R using the `lm` and `glm` functions.

This example is taken from the book *Extending the Linear Model with R*, by Julian J. Faraway. The book website, with code is

<http://www.maths.bath.ac.uk/~jjf23/ELM/>

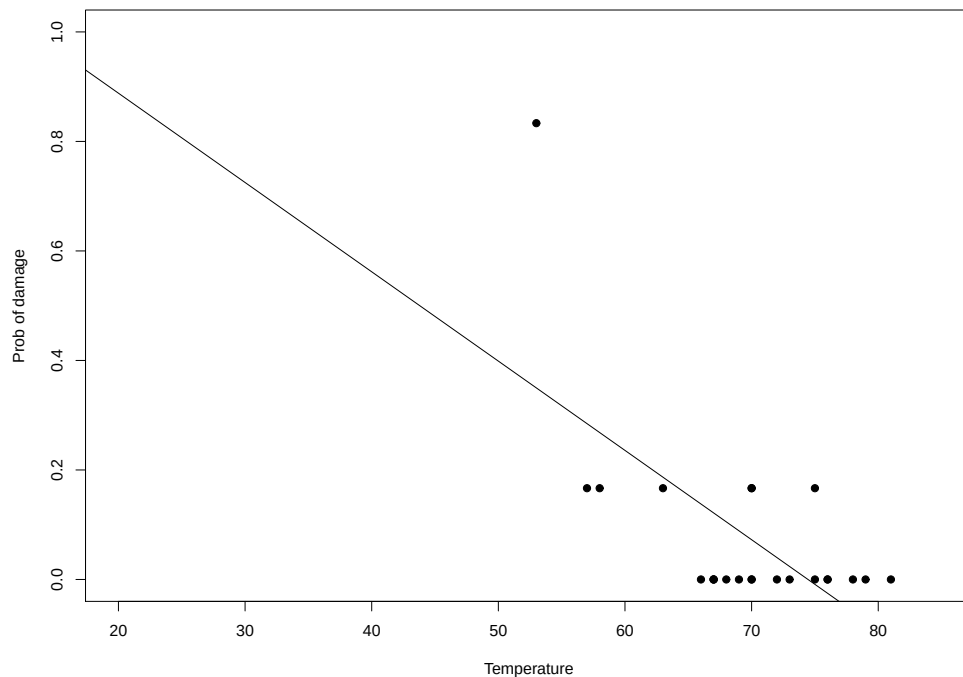


Figure 1: Linear Model Fit

EXAMPLE: CODE AND RESULTS

```

1 library(faraway)
2 data(orings)
3
4 plot(damage/6 ~ temp, orings, xlim=c(20,85), ylim = c(0,1),
5       xlab="Temperature", ylab="Prob of damage",pch=19)
6
7 lmod <- lm(damage/6 ~ temp, orings)
8 abline(lmod)
9
10 logitmod <- glm(cbind(damage,6-damage) ~ temp, family=binomial, orings)
11 plot(damage/6 ~ temp, orings, xlim=c(20,85), ylim = c(0,1),
12       xlab="Temperature", ylab="Prob of damage",pch=19)
13
14 x <- seq(20,85,1)
15 lines(x,ilogit(coef(logitmod)[1]+coef(logitmod)[2]*x))
16 probitmod <- glm(cbind(damage,6-damage) ~ temp,
17                  family=binomial(link=probit), orings)
18 lines(x,pnorm(coef(probitmod)[1]+coef(probitmod)[2]*x),lty=2)
19 abline(lmod,lty=3)
20 legend(20,0.4,c("Logit", "Probit", "Linear"),lty=c(1:3))
21
22 summary(lmod)
23 summary(logitmod)
24 summary(probitmod)
25
26 ilogit(coef(logitmod)[1]+coef(logitmod)[2]*31)

```

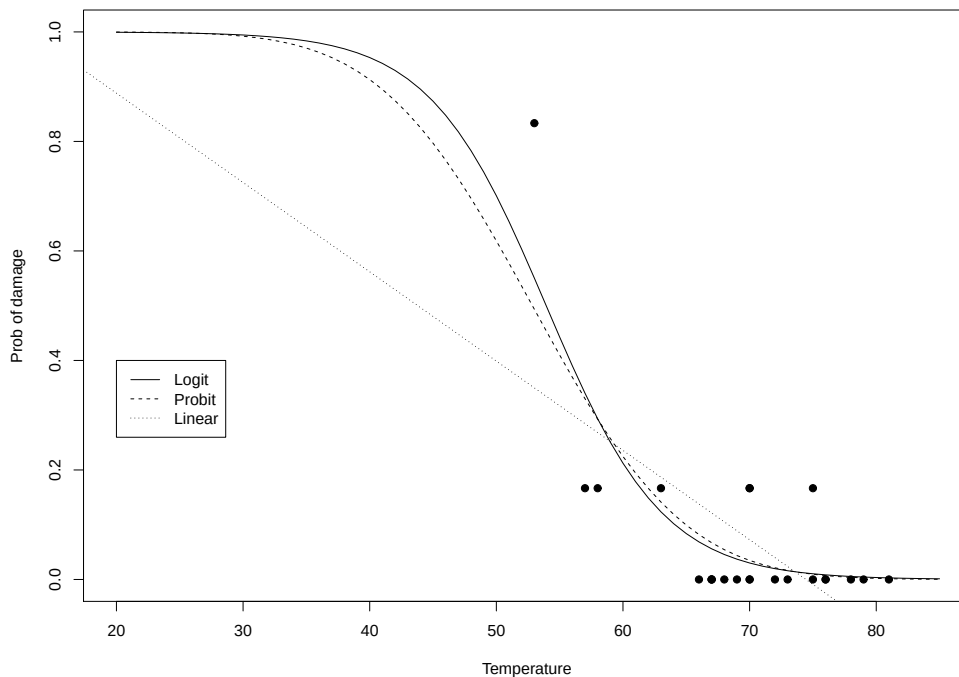


Figure 2: Linear, Logistic and Probit Model Fit

Results for the linear model:

```
> summary(lmod)
Call:
lm(formula = damage/6 ~ temp, data = orings)

Residuals:
    Min       1Q   Median       3Q      Max
-0.13786 -0.10345 -0.02369  0.06601  0.48345

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.21429     0.29993   4.049 0.000578 ***
temp        -0.01631     0.00429  -3.801 0.001043 **
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1

Residual standard error: 0.142 on 21 degrees of freedom
Multiple R-squared:  0.4076,    Adjusted R-squared:  0.3794
F-statistic: 14.45 on 1 and 21 DF,  p-value: 0.001043
```

Results for the logistic GLM:

```
> summary(logitmod)
Call:
glm(formula = cbind(damage, 6 - damage) ~ temp, family = binomial,
    data = orings)

Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-0.9529	-0.7345	-0.4393	-0.2079	1.9565

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	11.66299	3.29626	3.538	0.000403	***
temp	-0.21623	0.05318	-4.066	4.78e-05	***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 38.898 on 22 degrees of freedom
 Residual deviance: 16.912 on 21 degrees of freedom
 AIC: 33.675

Number of Fisher Scoring iterations: 6

Results for the probit GLM:

```
> summary(probitmod)
```

Call:

```
glm(formula = cbind(damage, 6 - damage) ~ temp, family = binomial(link = probit),
     data = orings)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0134	-0.7761	-0.4467	-0.1581	1.9983

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	5.59145	1.71055	3.269	0.00108	**
temp	-0.10580	0.02656	-3.984	6.79e-05	***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 38.898 on 22 degrees of freedom
 Residual deviance: 18.131 on 21 degrees of freedom
 AIC: 34.893

Number of Fisher Scoring iterations: 6

Comparison of fit at T=31:

```
> (coef(lmod)[1]+coef(lmod)[2]*31)
```

```
(Intercept)
```

```
0.7086905
```

```
> ilogit(coef(logitmod)[1]+coef(logitmod)[2]*31)
```

```
(Intercept)
```

```
0.9930342
```

```
> pnorm(coef(probitmod)[1]+coef(probitmod)[2]*31)
```

```
(Intercept)
```

```
0.9895983
```

2.4 ITERATIVELY REWEIGHTED LEAST SQUARES

In the Exponential-Dispersion family, with $a_i(\phi) = \phi/w_i$ for a set of known constants $w_1, \dots, w_n$, the p score equations for parameter $(\beta_1, \dots, \beta_p)^\top$ are obtained by differentiation of the log-likelihood partially with respect to the elements of β and equating to zero. For example, for β_j

$$\frac{\partial}{\partial \beta_j} \{l(\beta, \phi; \mathbf{y})\} = \sum_{i=1}^n \frac{w_i}{\phi} \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j} \right) = \sum_{i=1}^n \frac{w_i}{\phi} \left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i} \right) \frac{\partial \theta_i}{\partial \beta_j} = 0. \quad (1)$$

We have that

$$\mathbb{E}[Y_i] = \dot{b}(\theta_i) = \mu_i \quad \text{Var}[Y_i] = a_i(\phi) \ddot{b}(\theta_i) = \frac{\phi}{w_i} V(\mu_i)$$

say. After simplification, the score equations arising using link function $g(t)$ are, for $j = 1, \dots, p$

$$\sum_{i=1}^n \frac{w_i}{\phi} \frac{(y_i - \mu_i) x_{ij}}{\dot{g}(\mu_i) V(\mu_i)} = 0. \quad \text{or equivalently} \quad \sum_{i=1}^n \frac{w_i (y_i - \mu_i)}{V(\mu_i)} \frac{x_{ij}}{\dot{g}(\mu_i)} = 0 \quad (2)$$

Consider the random quantity $g(Y_i)$; by a Taylor approximation $g(Y_i) \simeq g(\mu_i) + (Y_i - \mu_i) \dot{g}(\mu_i)$, so let

$$Z_i = \eta_i + (Y_i - \mu_i) \dot{g}(\mu_i) = x_i \beta + (Y_i - \mu_i) \dot{g}(\mu_i) \quad (3)$$

As $\mathbb{E}[Y_i] = \mu_i$, we have that $\mathbb{E}[Z_i] = \eta_i$, and

$$\text{Var}[Z_i] = \text{Var}[Y_i] \{\dot{g}(\mu_i)\}^2 = \frac{\{\dot{g}(\mu_i)\}^2 \phi V(\mu_i)}{w_i} = v_i \quad (4)$$

say. Hence, from (3) the $\{Z_i\}$ form a **linear model** in β given $\{\mu_i\}$, and if observed values $\{z_i\}$ were available, β could be estimated using a **least squares** procedure, that is, by $\hat{\beta}$ where

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{z} \quad (5)$$

and where $\mathbf{X}$ is the $n \times n$ matrix whose n rows are $x_1, \dots, x_n$, with $x_i = (x_{i1}, \dots, x_{ip})$ and from (4)

$$\mathbf{W} = \text{diag}(v_1^{-1}, \dots, v_n^{-1})$$

ALGORITHM:

1 **Initialize:** choose any values for $\hat{\mu}_i^{(0)}$ and $\hat{\eta}_i^{(0)} = g(\hat{\mu}_i^{(0)})$ for $i = 1, \dots, n$

2 For $i = 1, \dots, n$ set

$$\hat{z}_i^{(1)} = \hat{\eta}_i^{(0)} + (y_i - \hat{\mu}_i^{(0)}) \dot{g}(\hat{\mu}_i^{(0)})$$

and

$$\hat{w}_i^{(1)} = \frac{w_i}{\{\dot{g}(\hat{\mu}_i^{(0)})\}^2 V(\hat{\mu}_i^{(0)})} \implies \mathbf{W}^{(1)} = \text{diag}(\hat{w}_1^{(1)}, \dots, \hat{w}_n^{(1)})$$

3 Compute estimates

$$\hat{\beta}^{(1)} = (\mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{z}^{(1)}$$

4 Define vectors $\hat{\eta}^{(1)} = \mathbf{X} \hat{\beta}^{(1)}$ and $\hat{\mu}^{(1)} = g^{-1}(\hat{\eta}^{(1)})$

5 Return to 2 and iterate through 2-5 using $\hat{\mu}^{(1)}$ and $\hat{\eta}^{(1)}$ as the updated starting values, eventually to obtain $\hat{\mu}^{(2)}$ and $\hat{\eta}^{(2)}$

6 Repeat until $\hat{\mu}^{(t)}$ and $\hat{\eta}^{(t)}$ satisfy

$$\left| \hat{\mu}^{(t)} - \hat{\mu}^{(t-1)} \right| < \epsilon_\mu \quad \left| \hat{\eta}^{(t)} - \hat{\eta}^{(t-1)} \right| < \epsilon_\eta$$

for **tolerances** ϵ_μ and ϵ_η .

This algorithm is known as **ITERATIVELY REWEIGHTED LEAST SQUARES**. It can be shown that, as $n \rightarrow \infty$,

$$\hat{\beta} \rightarrow N_p \left(\beta, \mathcal{I}(\beta)^{-1} \right) \quad \mathcal{I}(\beta) = \frac{1}{\phi} \mathbf{X}^\top \mathbf{W} \mathbf{X} \quad [\mathbf{W}]_{ii} = \frac{w_i}{V(\mu_i) \{\dot{g}(\mu_i)\}^2}$$

Example:- Poisson log-linear model.

Suppose $Y_i \sim \text{Poisson}(\mu_i)$, and $g(\mu_i) = \log \mu_i = \beta_0 + \beta_1 x_i$. Here $V(\mu_i) = \mu_i$, $\phi = 1$ and $w_i = 1$ for all i . Also note that $\dot{g}(t) = 1/t$, so that $V(\mu_i) \dot{g}(\mu_i) = \mu_i / \mu_i = 1$. We write by the Taylor approximation in (3)

$$z_i = (\beta_0 + \beta_1 x_i) + \frac{(y_i - \mu_i)}{\mu_i}$$

Step 0: Set $\hat{\beta}_0^{(0)} = \hat{\beta}_1^{(0)} = 0$, so that $\hat{\eta}_i^{(0)} = 0$, $\hat{\mu}_i^{(0)} = 1$, for all i .

Step 1: For $i = 1, \dots, n$, let

$$\hat{z}_i^{(1)} = \hat{\eta}_i^{(0)} + \frac{(y_i - \hat{\mu}_i^{(0)})}{\hat{\mu}_i^{(0)}} = 0 + \frac{(y_i - 1)}{1} = y_i - 1,$$

and

$$\hat{w}_i^{(1)} = \frac{w_i}{\{\dot{g}(\hat{\mu}_i^{(0)})\}^2 V(\hat{\mu}_i^{(0)})} = \frac{1}{\{1/\hat{\mu}_i^{(0)}\}^2 \hat{\mu}_i^{(0)}} = \hat{\mu}_i^{(0)} = 1$$

thus defining $\mathbf{W}^{(1)} = \text{diag}(\hat{w}_1^{(1)}, \dots, \hat{w}_n^{(1)}) = \mathbf{I}_n$. Then set

$$\hat{\beta}^{(1)} = \left(\mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{z}^{(1)} = \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{z}^{(1)}$$

and

$$\hat{\eta}_i^{(1)} = \hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)} x_i \quad \hat{\mu}_i^{(1)} = \exp\{\hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)} x_i\}$$

Step t+1: For $i = 1, \dots, n$, let

$$\hat{z}_i^{(t+1)} = \hat{\eta}_i^{(t)} + \frac{(y_i - \hat{\mu}_i^{(t)})}{\hat{\mu}_i^{(t)}} = (\hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} x_i) + \frac{(y_i - \hat{\mu}_i^{(t)})}{\hat{\mu}_i^{(t)}}$$

and

$$\hat{w}_i^{(t+1)} = \hat{\mu}_i^{(t)}$$

thus defining $\mathbf{W}^{(t+1)} = \text{diag}(\hat{w}_1^{(t+1)}, \dots, \hat{w}_n^{(t+1)})$. Then set

$$\hat{\beta}^{(t+1)} = \left(\mathbf{X}^\top \mathbf{W}^{(t+1)} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W}^{(t+1)} \mathbf{z}^{(t+1)}$$

Iterate Repeat until convergence; check that the score equations from (2)

$$\sum_{i=1}^n (y_i - \hat{\mu}_i^{(t)}) = 0 \quad \sum_{i=1}^n (y_i - \hat{\mu}_i^{(t)}) x_i = 0$$

holds at the solution.

ITERATIVELY REWEIGHTED LEAST SQUARES: CODE

```
1 set.seed(322)
2
3 n<-100
4 X<-runif(n,-10,10)
5 b0<-1.0
6 b1<-0.5
7 mu<-exp(b0+b1*X)
8 Y<-rpois(n,mu)
9
10 glm.p<-glm(Y~X,family=poisson)
11
12 #Initialization
13 Xmat<-cbind(rep(1,n),X)
14 beta<-c(0,0)
15 eta<-Xmat %*% beta
16 muvec<-exp(eta)
17 W<-diag(as.vector(muvec))
18 Z<-eta+(Y-muvec)/muvec
19
20 eqns<-t(Xmat) %*% (Y-muvec)
21
22 #Recursion
23
24 while(sum(eqns^2) > 1e-8){
25
26     beta<-solve(t(Xmat) %*% W %*% Xmat) %*% t(Xmat) %*% W %*% Z
27     eta<-Xmat %*% beta
28     muvec<-exp(eta)
29     W<-diag(as.vector(muvec))
30     Z<-eta+(Y-muvec)/muvec
31     eqns<-t(Xmat) %*% (Y-muvec)
32 }
33
34 print(istep)
35 print(t(beta))
36 print(sqrt(diag(solve(t(Xmat) %*% W %*% Xmat))))
37
38 #Compare results
39 summary(glm.p)
```

RESULTS

```
1 > print(istep)
2 [1] 155
3 > print(t(beta))
4           X
5 [1,] 1.058942 0.4941272
6 > print(sqrt(diag(solve(t(Xmat) %*% W %*% Xmat))))
7           X
8 0.06552041 0.00781589
9 >
10 > #Compare results
11 > summary(glm.p)
12
13 Call:
14 glm(formula = Y ~ X, family = poisson)
15
16 Deviance Residuals:
17      Min       1Q   Median       3Q      Max
18 -2.7779  -0.6767  -0.2606   0.5064   2.6571
19
20 Coefficients:
21             Estimate Std. Error z value Pr(>|z|)
22 (Intercept)  1.058942   0.065520   16.16  <2e-16 ***
23 X            0.494127   0.007816   63.22  <2e-16 ***
24 ---
25 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
26
27 (Dispersion parameter for poisson family taken to be 1)
28
29 Null deviance: 12184.02  on 99  degrees of freedom
30 Residual deviance:  109.39  on 98  degrees of freedom
31 AIC: 448.62
32
33 Number of Fisher Scoring iterations: 4
```

2.6 ASYMPTOTIC RESULTS FOR MAXIMUM LIKELIHOOD ESTIMATORS

In a probability model with independent data $y_1, \dots, y_n$, maximum likelihood estimation proceeds by solving the likelihood equations, obtained by setting the first derivative of the log-likelihood with respect to the parameters equal to zero. If $\theta = (\theta_1, \dots, \theta_p)^\top$ is the parameter vector, and

$$l_n(\theta) = l(\theta; \mathbf{y}) = \sum_{i=1}^n l(\theta; y_i) \quad (6)$$

is the log-likelihood, then

$$\dot{l}_n(\theta) = \dot{l}(\theta; \mathbf{y}) = \sum_{i=1}^n \dot{l}(\theta; y_i) = \sum_{i=1}^n \frac{\dot{f}_Y(y_i; \theta)}{f_Y(y_i; \theta)} = \sum_{i=1}^n S(y_i; \theta_0) \quad (7)$$

say, where $\dot{l}(\theta; y_i)$ is the $(p \times 1)$ vector of first partial derivatives, with j th element

$$\frac{\partial l(\theta; y_i)}{\partial \theta_j}$$

The function $\dot{l}(\theta; y_i) = S(y_i; \theta_0)$ is referred to as the **score function**. To compute the estimates, we must solve the $p \times 1$ system of **likelihood equations**

$$\dot{l}_n(\theta) = \dot{l}_n(\theta; \mathbf{y}) = 0_p \quad (\text{LE})$$

To understand the statistical properties of the resulting estimates and estimators, we consider the sampling distribution of the estimators assuming that the data are generated from the true distribution with parameter θ_0 , denoted $f_Y(y; \theta_0)$.

Recall that, under standard regularity conditions, if $Y_i \sim f_Y(y; \theta_0)$, then

$$\mathbb{E}_{f_Y}[\dot{l}(\theta_0; Y_i)] = \mathbb{E}_{f_Y}[S(Y_i; \theta_0)] = 0$$

Also, denote by $\mathcal{I}(\theta_0)$ the $p \times p$ matrix

$$\mathbb{E}_{f_Y}[S(Y_i; \theta_0)S(Y_i; \theta_0)^\top]$$

which can be shown to equal

$$-\mathbb{E}_{f_Y}[\Psi(Y_i; \theta_0)]$$

say, where $\Psi(Y_i; \theta_0)$ is the matrix of second derivatives with (j, k) th element given by

$$\frac{\partial^2 l(\theta; y)}{\partial \theta_j \partial \theta_k} \quad j, k = 1, \dots, p.$$

evaluated at $\theta = \theta_0$. This matrix is symmetric, and under regularity conditions, positive definite. The matrix $\mathcal{I}(\theta_0)$ is termed the **Fisher Information**.

Finally, recall the following distributional results: first, if $Y \sim \mathcal{N}(0, \sigma^2)$, then $(Y/\sigma)^2 \sim \chi_1^2$. Secondly, the multivariate equivalent of this result: if $Y \sim \mathcal{N}_p(0, \Sigma)$, then

$$Y^\top \Sigma^{-1} Y \sim \chi_p^2.$$

Consistency and Asymptotic Normality: Univariate Case

In the case $d = 1$, it is quite straightforward to show that a solution, $\hat{\theta}_n$, to the equation (LE) is *weakly consistent* for θ_0 , and *asymptotically normally distributed* provided that the log-likelihood is suitably differentiable with respect to θ on parameter space Θ , and otherwise regular.

- **Consistency:** As $n \rightarrow \infty$, we have that

$$\hat{\theta}_n \xrightarrow{p} \theta_0$$

- **Asymptotic Normality:** Consider a Taylor expansion of $\dot{l}_n(\theta)$ around θ_0

$$\dot{l}_n(\theta) = \dot{l}_n(\theta_0) + (\theta - \theta_0)\ddot{l}_n(\theta_0) + (\theta - \theta_0)^2\ddot{\ddot{l}}_n(\theta^*)/2$$

where θ^* lies between θ_0 and θ . Evaluating this at $\theta = \hat{\theta}_n$, a root of the likelihood equation

$$0 = \dot{l}_n(\hat{\theta}_n) = \dot{l}_n(\theta_0) + (\hat{\theta}_n - \theta_0)\ddot{l}_n(\theta_0) + (\hat{\theta}_n - \theta_0)^2\ddot{\ddot{l}}_n(\theta_n^*)/2$$

so that on rearrangement

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\dot{l}_n(\theta_0)/\sqrt{n}}{-(1/n)\ddot{l}_n(\theta_0) - (1/2n)(\hat{\theta}_n - \theta_0)\ddot{\ddot{l}}_n(\theta_n^*)}$$

Now, if

$$V(\theta_0) = \text{Var}_{f_Y}[S(Y; \theta_0)] = \mathcal{I}(\theta_0),$$

defines the **Fisher Information** $\mathcal{I}(\theta_0)$, then as $n \rightarrow \infty$, by the **Central Limit Theorem**,

$$\frac{1}{\sqrt{n}} \dot{l}_n(\theta_0) = \sqrt{n} \frac{1}{n} \left\{ \sum_{i=1}^n \frac{\dot{f}_Y(Y_i; \theta_0)}{f_Y(Y_i; \theta_0)} \right\} \xrightarrow{d} Z \sim \mathcal{N}(0, V(\theta_0))$$

Similarly, by the Weak Law of Large Numbers, as $n \rightarrow \infty$,

$$-\frac{1}{n} \ddot{l}_n(\theta_0) = -\frac{1}{n} \sum_{i=1}^n \ddot{l}(Y_i; \theta_0) \xrightarrow{p} \mathcal{I}(\theta_0).$$

Finally, under the assumption that the third derivative $\ddot{\ddot{l}}_n(\theta)$ is bounded in expectation, we have

$$\frac{1}{n}(\hat{\theta}_n - \theta_0)\ddot{\ddot{l}}_n(\theta_n^*) \xrightarrow{p} 0 \quad \therefore \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$$

These results extend easily to the vector parameter case: if θ is a p -dimensional vector parameter, then the maximum likelihood estimate is still weakly consistent, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}_p(0_p, \mathcal{I}(\theta_0)^{-1})$$

Observed Information

For practical purposes, it is often necessary to produce an estimator of $\mathcal{I}(\theta_0)$. The *observed information*, $\hat{I}_n = \hat{I}_n(\mathbf{Y})$, can be used to approximate $\mathcal{I}(\theta_0)$, where any of the following definitions may be used.

$$\hat{I}_n = \begin{cases} \mathcal{I}(\hat{\theta}_n) \\ \frac{1}{n} \sum_{i=1}^n S(y_i; \hat{\theta}_n) S(y_i; \hat{\theta}_n)^\top \\ -\frac{1}{n} \sum_{i=1}^n \Psi(y_i; \hat{\theta}_n) \end{cases}$$

THE LIKELIHOOD RATIO TEST

Recall that the **Likelihood Ratio Test (LRT)** statistic for testing H_0 against H_1

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

is based on the statistic

$$\lambda(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} f_{\mathbf{Y}}(\mathbf{y}; \theta)}{\sup_{\theta \in \Theta} f_{\mathbf{Y}}(\mathbf{y}; \theta)} = \frac{L(\hat{\theta}_{n0}; \mathbf{y})}{L(\hat{\theta}_n; \mathbf{y})}$$

say, where $\Theta \equiv \Theta_0 \cup \Theta_1$; H_0 is **rejected** if $\lambda(\mathbf{y})$ is **small enough**, that is, $\lambda(\mathbf{y}) \leq k$ for some k to be defined. **Note that Θ is not necessarily the entire parameter space, just the union of Θ_0 and Θ_1 .**

Asymptotic Behaviour of the Likelihood Ratio Test Statistic under the Null Hypothesis

Theorem Consider using the likelihood ratio statistic $\lambda(\mathbf{Y})$ for testing

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned} \tag{8}$$

Then under regularity assumptions, the maximum likelihood estimator for θ under H_1 , $\hat{\theta}_n$, is consistent and asymptotically normally distributed, and as $n \rightarrow \infty$ under H_0 ,

$$-2 \log \lambda(\mathbf{Y}) = -2(l_n(\theta_0) - l_n(\hat{\theta}_n)) \xrightarrow{d} Q \sim \chi_1^2$$

Proof A second-order Taylor expansion to the log-likelihood around the MLE, $\hat{\theta}_n$

$$l_n(\theta) = l_n(\hat{\theta}_n) + (\theta - \hat{\theta}_n) \dot{l}_n(\hat{\theta}_n) + \frac{1}{2}(\theta - \hat{\theta}_n)^2 \ddot{l}_n(\hat{\theta}_n) + \frac{1}{6}(\theta - \hat{\theta}_n)^3 \dddot{l}_n(\theta_n^*)$$

where θ_n^* lies between θ and $\hat{\theta}_n$. Now as $\hat{\theta}_n$ is the maximum likelihood estimate $\dot{l}_n(\hat{\theta}_n) = 0$, and therefore on rearrangement, evaluating at $\theta = \theta_0$,

$$-2(l_n(\theta_0) - l_n(\hat{\theta}_n)) = -(\theta_0 - \hat{\theta}_n)^2 \ddot{l}_n(\hat{\theta}_n) - \frac{1}{3}(\theta_0 - \hat{\theta}_n)^3 \dddot{l}_n(\theta_n^*).$$

In this expression, consider the asymptotic behaviour in terms of the random variables. We have

$$-(\theta_0 - \hat{\theta}_n)^2 \ddot{l}_n(\hat{\theta}_n) = (\sqrt{n}(\theta_0 - \hat{\theta}_n))^2 \left(-\frac{1}{n} \ddot{l}_n(\hat{\theta}_n) \right)$$

where

$$\sqrt{n}(\theta_0 - \hat{\theta}_n) \xrightarrow{d} Z \sim \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1}) \quad -\frac{1}{n} \ddot{l}_n(\hat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta_0)$$

so that, by a standard theorem (**Slutsky's Theorem**)

$$-(\theta_0 - \hat{\theta}_n)^2 \ddot{l}_n(\hat{\theta}_n) \xrightarrow{d} \chi_1^2$$

Finally,

$$(\theta_0 - \hat{\theta}_n)^3 \dddot{l}_n(\theta_n^*) = (\theta_0 - \hat{\theta}_n) \left((\theta_0 - \hat{\theta}_n)^2 \ddot{l}_n(\theta_n^*) \right) \xrightarrow{p} 0$$

so

$$-2(l_n(\theta_0) - l_n(\hat{\theta}_n)) \xrightarrow{d} Q \sim \chi_1^2$$

p-dimensional parameters

Consider using the likelihood ratio statistic for testing H_0 versus H_1 if θ is p -dimensional parameter. Under regularity conditions, as $n \rightarrow \infty$ under H_0 ,

$$-2 \log \lambda(\mathbf{Y}) = -2(l_n(\theta_0) - l_n(\hat{\theta}_n)) \xrightarrow{d} Q \sim \chi_p^2$$

This follows using identical methods to the $p = 1$ case. A second-order Taylor expansion to the log-likelihood around the MLE, $\hat{\theta}_n$, yields

$$l_n(\theta) \simeq l_n(\hat{\theta}_n) + (\theta - \hat{\theta}_n)^\top \dot{l}_n(\hat{\theta}_n) + \frac{1}{2}(\theta - \hat{\theta}_n)^\top \ddot{l}_n(\hat{\theta}_n)(\theta - \hat{\theta}_n)$$

As $\hat{\theta}_n$ is the maximum likelihood estimate

$$\dot{l}_n(\hat{\theta}_n) = 0_p$$

and therefore on rearrangement, evaluating at $\theta = \theta_0$,

$$-2(l_n(\theta_0) - l_n(\hat{\theta}_n)) \simeq -(\theta - \hat{\theta}_n)^\top \ddot{l}_n(\hat{\theta}_n)(\theta - \hat{\theta}_n)$$

But, by previous results

$$(\theta - \hat{\theta}_n)^\top \ddot{l}_n(\hat{\theta}_n)(\theta - \hat{\theta}_n) = \sqrt{n}(\theta - \hat{\theta}_n)^\top \left(-\frac{1}{n} \ddot{l}_n(\hat{\theta}_n) \right) \sqrt{n}(\theta - \hat{\theta}_n) \xrightarrow{d} \mathbf{Z}^\top \mathcal{I}(\theta_0) \mathbf{Z}$$

as

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathbf{Z} \sim \mathcal{N}_p(0_p, \mathcal{I}(\theta_0)^{-1}) \quad -\frac{1}{n} \ddot{l}_n(\hat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta_0)$$

Hence, by standard distributional results, as $\mathbf{Z}^\top \mathcal{I}(\theta_0) \mathbf{Z} \stackrel{d}{=} \mathbf{U}^\top \mathbf{U}$, where $\mathbf{U} \sim \mathcal{N}_p(0_p, \mathbf{I}_p)$, we have

$$-2(l_n(\theta_0) - l_n(\hat{\theta}_n)) \xrightarrow{d} Q \sim \chi_p^2.$$

Composite Hypotheses

Theorem Consider testing the hypothesis

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

using the likelihood ratio statistic

$$\lambda(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} f_{\mathbf{Y}|\theta}(\mathbf{y}; \theta)}{\sup_{\theta \in \Theta} f_{\mathbf{Y}|\theta}(\mathbf{y}; \theta)} = \frac{L(\hat{\theta}_{n0}; \mathbf{y})}{L(\hat{\theta}_n; \mathbf{y})}$$

say, where $\Theta \equiv \Theta_0 \cup \Theta_1$, and where H_0 specifies a model with k_1 free parameters (parameters not determined by the hypothesis), and H_1 specifies a model with k_2 free parameters, with $k_2 > k_1$.

Then, under regularity conditions, as $n \rightarrow \infty$, under H_0

$$-2 \log \lambda(\mathbf{Y}) = -2(l_n(\hat{\theta}_{n0}) - l_n(\hat{\theta}_n)) \xrightarrow{d} Q \sim \chi_{k_2 - k_1}^2$$

Note: Such hypotheses can often be specified in the form

$$\begin{aligned} H_0 &: \theta = (\theta_0, \theta_1), \quad \theta_1 \text{ unspecified} \\ H_1 &: \theta \neq (\theta_0, \theta_1), \quad \theta_0, \theta_1 \text{ unspecified} \end{aligned}$$

that is, H_0 places constraints on one component θ , but leaves the other unspecified.

2.7 RESIDUALS FOR GLMS

Three types of residual are used for checking the fit of a GLM

- **Pearson residual**

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

- **Anscombe residual**

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{\dot{A}(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}}$$

where function A is defined by

$$A(t) = \int_{-\infty}^t \frac{ds}{V^{1/3}(s)}$$

The function $A(\cdot)$ is chosen such that, for models in the Exponential/Exponential-Dispersion family, the skewness of the distribution of the transformed variable $A(Y)$ is as small as possible.

- **Deviance Residual**

$$D(\mathbf{y}; \hat{\theta}) = \frac{1}{\phi} \sum_{i=1}^n D_i(y_i; \hat{\theta}) = \frac{1}{\phi} \sum_{i=1}^n D_i$$

say, where

$$D_i = 2w_i \left\{ y_i(\tilde{\theta} - \hat{\theta}_i) - (b(\tilde{\theta}) - b(\hat{\theta}_i)) \right\}$$

Then the deviance residual is

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{D_i} = \begin{cases} \sqrt{D_i} & y_i > \hat{\mu}_i \\ -\sqrt{D_i} & y_i < \hat{\mu}_i \end{cases}$$

In the Poisson case, where $\hat{\mu}_i = \hat{\lambda}_i = \exp \{x_i \beta\}$,

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

$$r_{Ai} = \frac{3(y_i^{2/3} - \hat{\mu}_i^{2/3})}{2\hat{\mu}_i^{1/6}}$$

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i) \left\{ 2(y_i \log \frac{y_i}{\hat{\mu}_i} - y_i + \hat{\mu}_i) \right\}^{1/2}$$

CODE & RESULTS

```
1 f0<-glm(Y~1, family=poisson)
2 f1<-glm(Y~X, family=poisson)
3
4 rp0<-residuals(f0, type="pearson")
5 rd0<-residuals(f0, type="deviance")
6 fA0<-fitted(f0)
7 rA0<-1.5*(Y^(2/3)-fA0^(2/3))/fA0^(1/6)
8 rp1<-residuals(f1, type="pearson")
9 rd1<-residuals(f1, type="deviance")
10 fA1<-fitted(f1)
11 rA1<-1.5*(Y^(2/3)-fA1^(2/3))/fA1^(1/6)
```

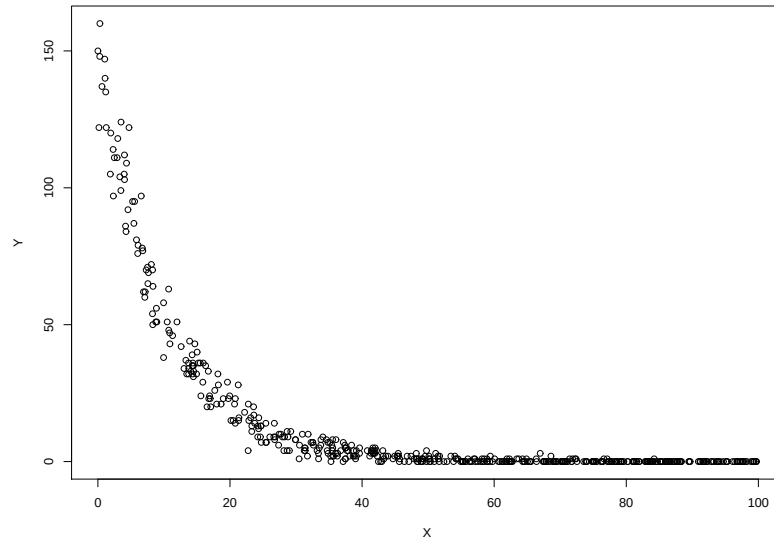


Figure 3: Poisson data Y with single covariate X

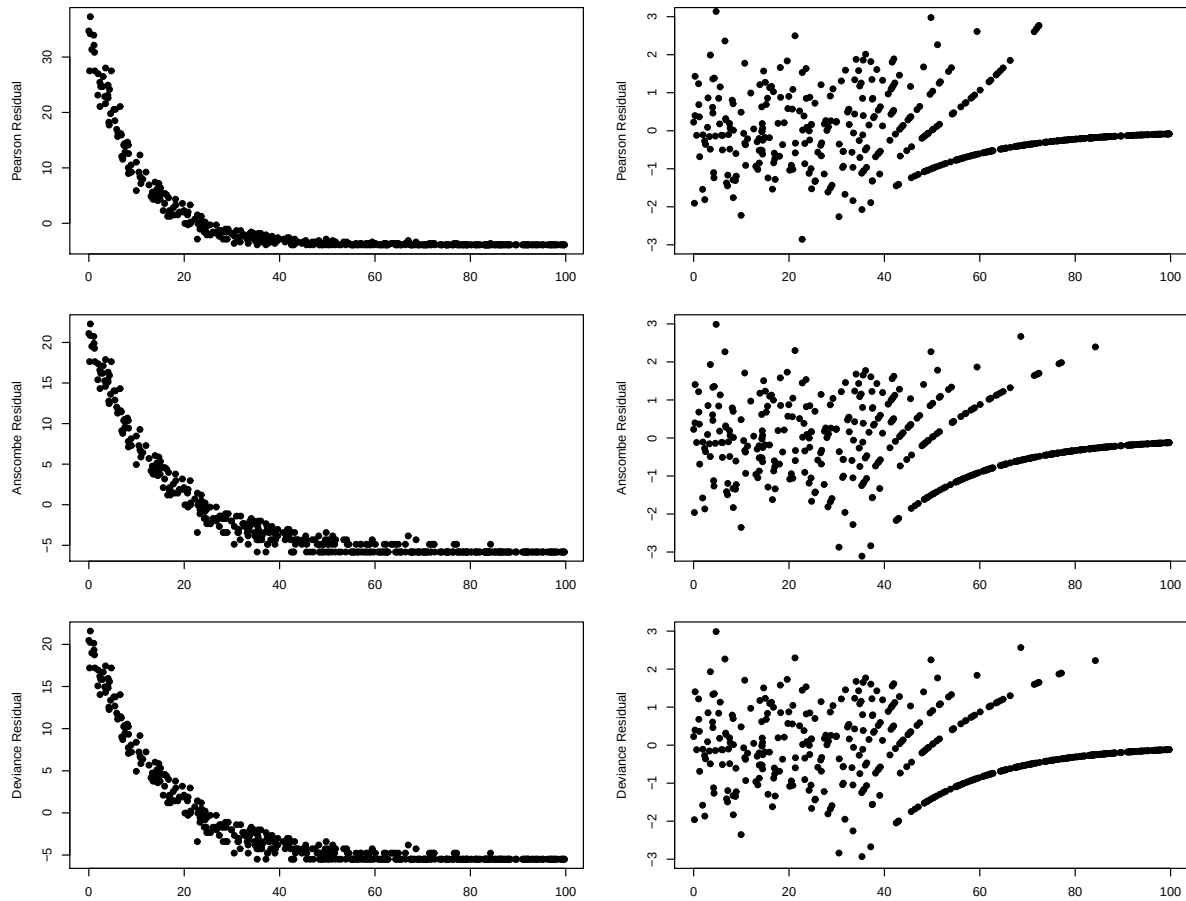


Figure 4: Residuals from fits of null (left column) and single predictor (right column): Pearson (top), Anscombe (middle) and deviance (bottom) residuals.

3 MODELS FOR COUNT DATA

EXAMPLE: POISSON DATA

An investigation of the incidence of the occurrences of four types of tumor at three different body locations involved recording counts from 400 randomly sampled cancer registry records.

- Tumor type: `type`
 - 1 Hutchinsons melanotic freckle
 - 2 Superficial spreading melanoma
 - 3 Nodular
 - 4 Indeterminate
- Location: `site`
 - 1 Head and Neck
 - 2 Trunk
 - 3 Extremities

		Site			Total
		1	2	3	
Tumor	1	22	2	10	34
	2	16	54	115	185
	3	19	33	73	125
	4	11	17	28	56
Total		68	106	226	400

A Poisson GLM can be used to analyze these data: the response data are the counts of numbers of tumors observed in each of the $4 \times 3 = 12$ cross-categorized groups defined by the levels of the two factors. Using a canonical log link function, four models can be fitted for the linear predictor,

- Model M_0 : Null model
- Model M_1 : Tumor type only main effect model
- Model M_2 : Site only main effect model
- Model M_3 : Tumor type plus Site additive main effects model

where all models are nested inside M_3 , where for $j = 1, \dots, 4, k = 1, \dots, 3$

$$\log \mu_{jk} = \log \mu + \alpha_j + \beta_k$$

with the usual identifiability constraints. These models are written

`1, tumor, site, tumor+site`

respectively. Fits of the four models are summarized in the following table: p is the total number of parameters in the fitted model, Dev. is residual deviance, and rdf is the residual degrees of freedom.

Model	p	Dev.	rdf
M_0	1	295.200	11
M_1	4	150.100	8
M_2	3	196.900	9
M_3	6	51.795	6

EXAMPLE: CODE & RESULTS

```
1 site<-gl(3,1,12)
2 tumor<-gl(4,3)
3 Y<-c(22,2,10,16,54,115,19,33,73,11,17,28)
4 cancer<-data.frame(tumor,site,Y)
5 cancer.m0 <- glm(Y ~ 1, family=poisson,data=cancer)
6 cancer.m1 <- glm(Y ~ tumor, family=poisson,data=cancer)
7 cancer.m2 <- glm(Y ~ site, family=poisson,data=cancer)
8 cancer.m3 <- glm(Y ~ tumor + site, family=poisson,data=cancer)
9
10 summary(cancer.m0)
11 summary(cancer.m1)
12 summary(cancer.m2)
13 summary(cancer.m3)
```

Note: The gl function generates levels of a factor.

```
14 > summary(cancer.m0)
15 Call: glm(formula = Y ~ 1, family = poisson, data = cancer)
16
17 Deviance Residuals:
18     Min       1Q   Median       3Q      Max
19 -7.170  -3.633  -2.398   0.777  11.022
20
21 Coefficients:
22             Estimate Std. Error z value Pr(>|z|)
23 (Intercept)    3.507      0.050   70.13  <2e-16 ***
24 ---
25 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
26 (Dispersion parameter for poisson family taken to be 1)
27
28     Null deviance: 295.20  on 11  degrees of freedom
29 Residual deviance: 295.20  on 11  degrees of freedom
30 AIC: 356.31
31
32 > summary(cancer.m1)
33 Call: glm(formula = Y ~ tumor, family = poisson, data = cancer)
34
35 Deviance Residuals:
36     Min       1Q   Median       3Q      Max
37 -6.940  -2.299  -0.701   2.208   6.055
38
39 Coefficients:
40             Estimate Std. Error z value Pr(>|z|)
41 (Intercept)    2.4277      0.1715  14.156  < 2e-16 ***
42 tumor2         1.6940      0.1866   9.079  < 2e-16 ***
43 tumor3         1.3020      0.1934   6.731 1.68e-11 ***
44 tumor4         0.4990      0.2174   2.295  0.0217 *
45 ---
46 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
47 (Dispersion parameter for poisson family taken to be 1)
48
49     Null deviance: 295.20  on 11  degrees of freedom
50 Residual deviance: 150.10  on  8  degrees of freedom
51 AIC: 217.21
52
```

```

53 > summary(cancer.m2)
54 Call:
55 glm(formula = Y ~ site, family = poisson, data = cancer)
56
57 Deviance Residuals:
58      Min       1Q   Median       3Q      Max
59 -7.6398  -2.5337   0.1155   1.4367   6.8161
60
61 Coefficients:
62             Estimate Std. Error z value Pr(>|z|)
63 (Intercept)    2.8332     0.1213  23.363 < 2e-16 ***
64 site2          0.4439     0.1554   2.857  0.00427 **
65 site3          1.2010     0.1383   8.683 < 2e-16 ***
66 ---
67 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
68
69 (Dispersion parameter for poisson family taken to be 1)
70
71 Null deviance: 295.20  on 11  degrees of freedom
72 Residual deviance: 196.90  on  9  degrees of freedom
73 AIC: 262.01
74
75 > summary(cancer.m3)
76 Call:glm(formula = Y ~ tumor + site, family = poisson, data = cancer)
77
78 Deviance Residuals:
79      Min       1Q   Median       3Q      Max
80 -3.0453  -1.0741   0.1297   0.5857   5.1354
81
82 Coefficients:
83             Estimate Std. Error z value Pr(>|z|)
84 (Intercept)    1.7544     0.2040   8.600 < 2e-16 ***
85 tumor2         1.6940     0.1866   9.079 < 2e-16 ***
86 tumor3         1.3020     0.1934   6.731 1.68e-11 ***
87 tumor4         0.4990     0.2174   2.295  0.02173 *
88 site2          0.4439     0.1554   2.857  0.00427 **
89 site3          1.2010     0.1383   8.683 < 2e-16 ***
90 ---
91 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
92 (Dispersion parameter for poisson family taken to be 1)
93
94 Null deviance: 295.203  on 11  degrees of freedom
95 Residual deviance:  51.795  on  6  degrees of freedom
96 AIC: 122.91

```

EXAMPLE: POISSON DATA WITH OFFSET

A Poisson GLM can also be used in situations where a binomial model might seem more appropriate. It is well known that if $X \sim \text{Binomial}(m, \pi)$, with m large and π small, then, approximately

$$X \sim \text{Poisson}(\mu)$$

with $\mu = m\pi$. Therefore, we might consider a GLM where

$$\log \mu_i = \log m_i + \log \pi_i$$

where π_i is considered a standardized Poisson rate. In this representation, however, as m_i is **fixed**, it must be included in the model with a fixed coefficient, equal to one; such a model term is referred to as an **offset**.

EXAMPLE: The numbers of lung cancer cases in four different smoking groups is recorded, and the variation in the numbers of cases across nine different age groups is inspected. In the study population, the partition into the 4×36 cross-classified subgroups involved a total of 56122 thousand individuals.

- Smoking group: `smoke`
 - 1 Does not smoke
 - 2 Smokes cigars or pipe only
 - 3 Smokes cigarettes and cigar or pipe
 - 4 Smokes cigarettes only
- Age Group: `age`
 - 1-9 for age groups 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+ respectively

		Age Group								
		40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
Smoking group	1 Dead	18	22	19	55	117	170	179	120	120
	Popn (thous.)	656	359	249	632	1067	897	668	361	274
	2 Dead	2	4	3	38	113	173	212	243	253
	Popn (thous.)	145	104	98	372	846	949	824	667	537
	3 Dead	149	169	193	576	1001	901	613	337	189
	Popn (thous.)	4531	3030	2267	4682	6052	3880	2033	871	345
	4 Dead	124	140	187	514	778	689	432	214	63
	Popn (thous.)	3410	2239	1851	3270	3791	2421	1195	436	113

Using a canonical log link function with offset, four models can be fitted for the linear predictor,

- Model M_0 : Null model
- Model M_1 : Smoking group only main effect model
- Model M_2 : Age group only main effect model
- Model M_3 : Smoking group plus age group additive main effects model

where all models are nested inside M_3 , where for $j = 1, \dots, 4, k = 1, \dots, 3$

$$\log \mu_{jk} = \log m_{jk} + \alpha_j + \beta_k$$

with the usual identifiability constraints. These models are written

$$1, \text{smoke}, \text{age}, \text{smoke} + \text{age}$$

Fits of the four models are summarized in the following table: p is the total number of parameters in the fitted model, Dev. is residual deviance, and rdf is the residual degrees of freedom.

Model	p	Dev.	rdf
M_0	1	4056.000	35
M_1	4	3910.700	32
M_2	9	191.720	27
M_3	12	21.487	24

On deviance assessment, it is clear that the model M_3 is the most sensible model for these data.

```

1 smoking<-read.table("smoking.dat",header=T)
2
3 smoking$smoke<-gl(4,9)
4
5 smoking.m0<-glm(dead~offset(log(pop)),family=poisson,data=smoking)
6 smoking.m1<-glm(dead~offset(log(pop))+smoke,family=poisson,data=smoking)
7 smoking.m2<-glm(dead~offset(log(pop))+age,family=poisson,data=smoking)
8 smoking.m3<-glm(dead~offset(log(pop))+smoke+age,family=poisson,data=smoking)
9
10 summary(smoking.m0)
11 summary(smoking.m1)
12 summary(smoking.m2)
13 summary(smoking.m3)

```

Results for Model M_3 :

Call:

```
glm(formula = dead ~ offset(log(pop)) + smoke + age, family = poisson,
    data = smoking)
```

Coefficients:

```

      Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.68002    0.06824 -53.929 < 2e-16 ***
smoke2       0.04781    0.04699   1.017   0.309
smoke3       0.21796    0.03869   5.633 1.77e-08 ***
smoke4       0.41696    0.03991  10.447 < 2e-16 ***
age45-59     0.55388    0.07999   6.924 4.38e-12 ***
age50-54     0.98039    0.07682  12.762 < 2e-16 ***
age55-59     1.37946    0.06526  21.138 < 2e-16 ***
age60-64     1.65423    0.06257  26.439 < 2e-16 ***
age65-69     1.99817    0.06279  31.824 < 2e-16 ***
age70-74     2.27141    0.06435  35.296 < 2e-16 ***
age75-79     2.55858    0.06778  37.746 < 2e-16 ***
age80+       2.84692    0.07242  39.310 < 2e-16 ***

```

```
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
```

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4055.984 on 35 degrees of freedom

Residual deviance: 21.487 on 24 degrees of freedom

AIC: 285.51

Number of Fisher Scoring iterations: 4

POISSON REGRESSION WITH COVARIATES AND FACTOR PREDICTORS

EXAMPLE: Scottish Lip Cancer data

District	Observed y	Expected E	%AFF X_1	Lat. X_2	Long. X_3
1	9	1.4	16	57.29	5.50
2	39	8.7	16	57.56	2.36
3	11	3.0	10	58.44	3.90
4	9	2.5	24	55.76	2.40
5	15	4.3	10	57.71	5.09
6	8	2.4	24	59.13	3.25
7	26	8.1	10	57.47	3.30
8	7	2.3	7	60.24	1.43
9	6	2.0	7	56.90	5.42
10	20	6.6	16	57.24	2.60
11	13	4.4	7	58.12	6.80
12	5	1.8	16	58.06	4.64
13	3	1.1	10	57.47	3.98
14	8	3.3	24	54.94	5.00
15	17	7.8	7	56.30	3.10
16	9	4.6	16	57.00	3.00
17	2	1.1	10	57.06	4.09
18	7	4.2	7	55.65	2.88
19	9	5.5	7	57.24	4.73
20	7	4.4	10	55.35	2.90
21	16	10.5	7	56.75	2.98
22	31	22.7	16	57.12	2.20
23	11	8.8	10	56.40	5.27
24	7	5.6	7	55.63	3.96
25	19	15.5	1	56.20	3.30
26	15	12.5	1	56.10	3.60
27	7	6.0	7	55.24	4.09
28	10	9.0	7	55.95	2.80
29	16	14.4	10	56.60	4.09
30	11	10.2	10	55.90	3.80
31	5	4.8	7	55.47	4.55
32	3	2.9	24	55.00	4.36
33	7	7.0	10	55.83	3.20
34	8	8.5	7	56.30	4.73
35	11	12.3	7	55.29	4.98
36	9	10.1	0	55.94	4.95
37	11	12.7	10	55.76	5.02
38	8	9.4	1	55.91	4.18
39	6	7.2	16	56.15	4.99
40	4	5.3	0	56.05	4.91
41	10	18.8	1	55.88	4.82
42	8	15.8	16	56.03	4.00
43	2	4.3	16	56.15	3.96
44	6	14.6	0	55.82	4.09
45	19	50.7	1	55.93	3.40
46	3	8.2	7	55.65	4.75
47	2	5.6	1	55.71	4.45
48	3	9.3	1	55.79	4.27
49	28	88.7	0	55.90	4.55
50	6	19.6	1	56.45	3.20
51	1	3.4	1	56.00	4.27
52	1	3.6	0	56.15	4.64
53	1	5.7	1	55.79	4.70
54	1	7.0	1	55.99	4.45
55	0	4.2	16	55.68	3.38
56	0	1.8	10	55.18	3.40

- Column 1: District number
- Column 2: Observed number of cases
- Column 3: Expected number of cases
- Column 4: Percent working in Agriculture, Fisheries, or Forestry (see text)
- Column 5: Latitude
- Column 6: Longitude

DATA BASED ON FACTOR PREDICTORS

TWO-WAY CONTINGENCY TABLES: Cross-classification by **two** factor predictors, no obvious response variable.

EXAMPLE: Political Affiliations of College Students: A 4×3 two-way table.

		Affiliation			Total
		Rep.	Dem.	Ind.	
College	Letters	34	61	16	111
	Engineering	31	19	17	67
	Agriculture	19	23	16	58
	Education	23	39	12	74
Total		107	142	61	310

- Factor 1: College Type (4 levels)
- Factor 2: Political affiliation (3 levels)

MULTI-WAY CONTINGENCY TABLES: Cross-classification by **multiple** factor predictors, no obvious response variable.

EXAMPLE: Classroom behaviour data: A $2 \times 2 \times 3$ multiway table.

		Adversity						
		Low		Med		High		
	Risk	N	Y	N	Y	N	Y	Total
Behaviour	Non-deviant	16	7	15	34	5	3	80
	Deviant	1	1	3	8	1	3	17
Total		17	8	18	42	6	6	97

- Factor 1: Classroom behaviour type: 1-Non-Deviant, 2-Deviant
- Factor 2: Risk at Home: Y-At Risk, N-Not at Risk
- Factor 3: Adversity of School: 1-Low, 2-Medium, 3-High.

ANALYSIS

EXAMPLE: Scottish Lip Cancer data

```
1 ScottishLip<-read.table("ScottishLipCancer.txt")
2 names(ScottishLip)<-c("ID","Y","E","AFF","LAT","LONG")
3
4 glm0<-glm(Y~1,family=poisson,data=ScottishLip)
5 glm0.off<-glm(Y~1+offset(log(E)),family=poisson,data=ScottishLip)
6 glm1.off<-glm(Y~AFF+offset(log(E)),family=poisson,data=ScottishLip)
7 glm2.off<-glm(Y~AFF+LAT+offset(log(E)),family=poisson,data=ScottishLip)
8 glm3.off<-glm(Y~AFF+LAT+LONG+offset(log(E)),family=poisson,data=ScottishLip)
9 glm4.off<-glm(Y~AFF+LAT+AFF:LAT+offset(log(E)),family=poisson,data=ScottishLip)
10 anova(glm0.off,glm1.off,glm2.off,glm3.off)
11 summary(glm2.off)
12 anova(glm0.off,glm1.off,glm2.off,glm4.off)
13 summary(glm4.off)
14 glm5.off<-glm(Y~AFF+LAT+AFF:LAT+offset(log(E)),
15               family=quasipoisson,data=ScottishLip)
16 summary(glm5.off)
```

Notes:

- An offset model is used, with offset equal to the Expected value.
- The `anova` function can be used to tabulate the Analysis of Deviance tests.
- The colon symbol in lines 9 and 14 indicates an interaction.
- The `family=quasipoisson` command in line 15 uses an estimate of dispersion ϕ , instead of setting $\phi = 1$ to compute the standard errors.

```
1 > anova(glm0.off,glm1.off,glm2.off,glm3.off)
2 Analysis of Deviance Table
3
4 Model 1: Y ~ 1 + offset(log(E))
5 Model 2: Y ~ AFF + offset(log(E))
6 Model 3: Y ~ AFF + LAT + offset(log(E))
7 Model 4: Y ~ AFF + LAT + LONG + offset(log(E))
8   Resid. Df Resid. Dev Df Deviance
9 1          55      380.73
10 2          54      238.62 1    142.106
11 3          53      159.71 1     78.908
12 4          52      159.44 1      0.276
13 >
14 > summary(glm2.off)
15
16 Call:
17 glm(formula = Y ~ AFF + LAT + offset(log(E)), family = poisson,
18     data = ScottishLip)
19
20 Deviance Residuals:
21      Min       1Q   Median       3Q      Max
22 -3.3979  -1.0943   0.2294   1.3201   3.0078
23
24 Coefficients:
25             Estimate Std. Error z value Pr(>|z|)
26 (Intercept) -24.586752    2.537728  -9.688  < 2e-16 ***
27 AFF          0.054241    0.006707   8.087  6.1e-16 ***
```

```

28 LAT          0.429060    0.045241    9.484 < 2e-16 ***
29 ---
30 Signif. codes:  0   ***    0.001   **    0.01   *    0.05   .    0.1     1
31
32 (Dispersion parameter for poisson family taken to be 1)
33
34 Null deviance: 380.73 on 55 degrees of freedom
35 Residual deviance: 159.71 on 53 degrees of freedom
36 AIC: 373.69
37
38 Number of Fisher Scoring iterations: 5
39
40 >
41 > anova(glm0.off,glm1.off,glm2.off,glm4.off)
42 Analysis of Deviance Table
43
44 Model 1: Y ~ 1 + offset(log(E))
45 Model 2: Y ~ AFF + offset(log(E))
46 Model 3: Y ~ AFF + LAT + offset(log(E))
47 Model 4: Y ~ AFF + LAT + AFF:LAT + offset(log(E))
48
49 Resid. Df Resid. Dev Df Deviance
50 1          55          380.73
51 2          54          238.62 1    142.106
52 3          53          159.71 1     78.908
53 4          52          152.86 1      6.848
54
55
56
57 > summary(glm4.off)
58
59 Call:
60 glm(formula = Y ~ AFF + LAT + AFF:LAT + offset(log(E)), family = poisson,
61     data = ScottishLip)
62
63 Deviance Residuals:
64     Min       1Q   Median       3Q      Max
65 -3.6309 -1.1398  0.1526  1.3215  3.2172
66
67 Coefficients:
68             Estimate Std. Error z value Pr(>|z|)
69 (Intercept) -34.594815    4.433805  -7.803 6.07e-15 ***
70 AFF          0.907269    0.333043   2.724 0.00645 **
71 LAT          0.606971    0.078845   7.698 1.38e-14 ***
72 AFF:LAT     -0.015117    0.005906  -2.560 0.01048 *
73 ---
74 Signif. codes:  0   ***    0.001   **    0.01   *    0.05   .    0.1     1
75
76 (Dispersion parameter for poisson family taken to be 1)
77
78 Null deviance: 380.73 on 55 degrees of freedom
79 Residual deviance: 152.86 on 52 degrees of freedom
80 AIC: 368.84
81
82 Number of Fisher Scoring iterations: 5

```

```

83
84 > glm5.off<-glm(Y~AFF+LAT+AFF:LAT+offset(log(E)), family=quasipoisson,
85               data=ScottishLip)
86 > summary(glm5.off)
87
88 Call:
89 glm(formula = Y ~ AFF + LAT + AFF:LAT + offset(log(E)), family = quasipoisson,
90     data = ScottishLip)
91
92 Deviance Residuals:
93     Min       1Q   Median       3Q      Max
94 -3.6309  -1.1398   0.1526   1.3215   3.2172
95
96 Coefficients:
97             Estimate Std. Error t value Pr(>|t|)
98 (Intercept) -34.59482     7.62767  -4.535 3.42e-05 ***
99 AFF          0.90727     0.57295   1.584  0.119
100 LAT          0.60697     0.13564   4.475 4.19e-05 ***
101 AFF:LAT      -0.01512     0.01016  -1.488  0.143
102 ---
103 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
104
105 (Dispersion parameter for quasipoisson family taken to be 2.959582)
106
107 Null deviance: 380.73  on 55  degrees of freedom
108 Residual deviance: 152.86  on 52  degrees of freedom
109 AIC: NA
110
111 Number of Fisher Scoring iterations: 5

```

EXAMPLE: Political Affiliations of College Students:

```

1 Y.aff<-c(34,61,16,31,19,17,19,23,16,23,39,12)
2 C.aff<-gl(4,3)
3 A.aff<-gl(3,1,12)
4
5 Affil<-data.frame(Y.aff,C.aff,A.aff)
6 names(Affil)<-c("Y","College","Affiliation")
7
8 glm0.aff<-glm(Y~1,family=poisson,data=Affil)
9 glm1.aff<-glm(Y~C,family=poisson,data=Affil)
10 glm2.aff<-glm(Y~A,family=poisson,data=Affil)
11 glm3.aff<-glm(Y~A+C,family=poisson,data=Affil)
12 glm4.aff<-glm(Y~A*C,family=poisson,data=Affil)
13
14 anova(glm3.aff,glm4.aff)
15 qchisq(0.95,6)
16
17 anova(glm2.aff,glm3.aff)
18 anova(glm1.aff,glm3.aff)
19
20 summary(glm4.aff)
21
22 fitted(glm4.aff)

```

Notes:

- The * in line 12 fits the *full factorial* model (that is, main effects and all interactions), which for the log-linear case is the full or *saturated* model.

```
1 > anova(glm3.aff,glm4.aff)
2 Analysis of Deviance Table
3
4 Model 1: Y ~ A + C
5 Model 2: Y ~ A * C
6   Resid. Df Resid. Dev Df Deviance
7 1          6      16.39
8 2          0          0.00  6    16.39
9
10 > qchisq(0.95,6)
11 [1] 12.59159
12
13 > anova(glm2.aff,glm3.aff)
14 Analysis of Deviance Table
15
16 Model 1: Y ~ A
17 Model 2: Y ~ A + C
18   Resid. Df Resid. Dev Df Deviance
19 1          9      36.175
20 2          6      16.390  3    19.785
21 > anova(glm1.aff,glm3.aff)
22 Analysis of Deviance Table
23
24 Model 1: Y ~ C
25 Model 2: Y ~ A + C
26   Resid. Df Resid. Dev Df Deviance
27 1          8      49.822
28 2          6      16.390  2    33.432
29
30 > summary(glm4.aff)
31
32 Call:
33 glm(formula = Y ~ A * C, family = poisson, data = Affil)
34
35
36           Estimate Std. Error z value Pr(>|z|)
37 (Intercept)  3.52636    0.17150  20.562 < 2e-16 ***
38 A2           0.58451    0.21402   2.731  0.00631 **
39 A3          -0.75377    0.30317  -2.486  0.01291 *
40 C2          -0.09237    0.24833  -0.372  0.70991
41 C3          -0.58192    0.28643  -2.032  0.04219 *
42 C4          -0.39087    0.26998  -1.448  0.14769
43 A2:C2       -1.07406    0.36152  -2.971  0.00297 **
44 A3:C2        0.15300    0.42778   0.358  0.72060
45 A2:C3       -0.39346    0.37672  -1.044  0.29628
46 A3:C3        0.58192    0.45502   1.279  0.20094
47 A2:C4       -0.05645    0.33901  -0.167  0.86776
48 A3:C4        0.10318    0.46768   0.221  0.82538
49 ---
50 Signif. codes:  0   ***    0.001   **   0.01   *   0.05   .   0.1   1
```

```

51 (Dispersion parameter for poisson family taken to be 1)
52
53 Null deviance: 6.9607e+01 on 11 degrees of freedom
54 Residual deviance: -2.2204e-15 on 0 degrees of freedom
55 AIC: 83.906 Number of Fisher Scoring iterations: 3
56
57 > fitted(glm4.aff)
58 1 2 3 4 5 6 7 8 9 10 11 12
59 34 61 16 31 19 17 19 23 16 23 39 12

```

EXAMPLE: Classroom behaviour data:

```

1 Y.behav<-c(16,7,15,34,5,3,1,1,3,8,1,3)
2 Y.Dev<-factor(rep(c("ND","D"),each=6))
3 Y.Risk<-factor(rep(c("N","Y"),6))
4 Y.School<-rep(factor(rep(c("Low","Med","High"),each=2)),2)
5
6 Behaviour<-data.frame(Y.behav,Y.Dev,Y.Risk,Y.School)
7 names(Behaviour)<-c("Y","Deviant","AtRisk","Adversity")
8
9 glm7.Behav<-glm(Y~Deviant*AtRisk*Adversity,family=poisson,data=Behaviour)
10
11 anova(glm7.Behav)
12
13 glm6.Behav<-update(glm7.Behav,~.-Deviant:AtRisk:Adversity)
14 summary(glm6.Behav)
15
16 glm5.Behav<-update(glm6.Behav,~.-AtRisk:Adversity)
17 summary(glm5.Behav)
18
19 anova(glm5.Behav,glm6.Behav,glm7.Behav)

```

Notes:

- The `anova` function in line 11 produces a sequence of fits of all the models nested inside the saturated model `glm7.Behav`
- The `update` function allows previously fitted models to be updated by the omission or inclusion of one or more terms.

```

1 >
2 > anova(glm7.Behav)
3 Analysis of Deviance Table
4
5 Model: poisson, link: log
6
7 Response: Y
8
9 Terms added sequentially (first to last)
10
11
12

```

	Df	Deviance	Resid. Df	Resid. Dev
13 NULL			11	100.718
14 Deviant	1	44.430	10	56.288
15 AtRisk	1	2.329	9	53.959
16 Adversity	2	37.540	7	16.419
17 Deviant:AtRisk	1	1.442	6	14.977

```

18 Deviant:Adversity          2      3.656          4      11.321
19 AtRisk:Adversity           2     10.378          2       0.943
20 Deviant:AtRisk:Adversity   2      0.943          0       0.000
21 >
22 > glm6.Behav<-update(glm7.Behav,~.-Deviant:AtRisk:Adversity)
23 > summary(glm6.Behav)
24
25 Call:
26 glm(formula = Y ~ Deviant + AtRisk + Adversity + Deviant:AtRisk +
27     Deviant:Adversity + AtRisk:Adversity, family = poisson, data = Behaviour)
28
29 Deviance Residuals:
30      1      2      3      4      5      6      7
31  0.02682 -0.04026 -0.18240  0.12394  0.28502 -0.33107 -0.10345  0.11113
32  0.45288 -0.24649 -0.51762
33  0.37972
34
35 Coefficients:
36             Estimate Std. Error z value Pr(>|z|)
37 (Intercept)      0.4766    0.6576   0.725   0.4686
38 DeviantND        1.0026    0.7046   1.423   0.1548
39 AtRiskY          0.3945    0.7150   0.552   0.5811
40 AdversityLow     -0.3749    0.9530  -0.393   0.6940
41 AdversityMed      0.3487    0.7357   0.474   0.6355
42 DeviantND:AtRiskY -0.5899    0.6151  -0.959   0.3375
43 DeviantND:AdversityLow  1.6615    0.9666   1.719   0.0856 .
44 DeviantND:AdversityMed  0.9269    0.7169   1.293   0.1961
45 AtRiskY:AdversityLow  -0.6094    0.7388  -0.825   0.4094
46 AtRiskY:AdversityMed   0.9456    0.6576   1.438   0.1504
47 ---
48 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
49
50 (Dispersion parameter for poisson family taken to be 1)
51
52 Null deviance: 100.71803 on 11 degrees of freedom
53 Residual deviance:  0.94285 on  2 degrees of freedom
54 AIC: 61.692
55
56 Number of Fisher Scoring iterations: 4
57
58 >
59 > glm5.Behav<-update(glm6.Behav,~.-AtRisk:Adversity)
60 > summary(glm5.Behav)
61
62 Call:
63 glm(formula = Y ~ Deviant + AtRisk + Adversity + Deviant:AtRisk +
64     Deviant:Adversity, family = poisson, data = Behaviour)
65
66 Deviance Residuals:
67      1      2      3      4      5      6      7
68  1.62458 -1.73653 -1.59441  1.30447  0.69645 -0.70855  0.48757 -0.36585
69 -0.13245  0.08402 -0.16704

```

```

69      12
70 0.10395
71
72 Coefficients:
73             Estimate Std. Error z value Pr(>|z|)
74 (Intercept)      0.1625     0.6254   0.260   0.7950
75 DeviantND        1.1184     0.7290   1.534   0.1250
76 AtRiskY          0.8755     0.5323   1.645   0.1000
77 AdversityLow     -0.6931     0.8660  -0.800   0.4235
78 AdversityMed      1.0116     0.5839   1.733   0.0832 .
79 DeviantND:AtRiskY -0.6748     0.5778  -1.168   0.2428
80 DeviantND:AdversityLow 1.7492     0.9584   1.825   0.0680 .
81 DeviantND:AdversityMed 0.8008     0.6974   1.148   0.2508
82 ---
83 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
84
85 (Dispersion parameter for poisson family taken to be 1)
86
87 Null deviance: 100.718  on 11  degrees of freedom
88 Residual deviance: 11.321  on  4  degrees of freedom
89 AIC: 68.07
90
91 Number of Fisher Scoring iterations: 4
92
93 >
94 > anova(glm5.Behav,glm6.Behav,glm7.Behav)
95 Analysis of Deviance Table
96
97 Model 1: Y ~ Deviant + AtRisk + Adversity + Deviant:AtRisk + Deviant:Adversity
98 Model 2: Y ~ Deviant + AtRisk + Adversity + Deviant:AtRisk + Deviant:Adversity +
99   AtRisk:Adversity
100 Model 3: Y ~ Deviant * AtRisk * Adversity
101
102   Resid. Df Resid. Dev Df Deviance
103 1         4      11.3206
104 2         2       0.9428 2  10.3777
105 3         0       0.0000 2   0.9428

```

DEVIANCE COMPARISON WITH ESTIMATED DISPERSION

When ϕ is known, the Analysis of Deviance for nested models $M_0 \subset M_1$ containing $p_0 < p_1$ parameters respectively, with deviances D_0 and D_1 , is based on the asymptotic theory of likelihood ratio tests: we have, for GLM with dispersion ϕ .

$$\frac{D_0 - D_1}{\phi} \rightsquigarrow \chi_{p_1 - p_0}^2$$

We also have the result that if the model M (possibly, but not necessarily, model M_1) with p parameters is an adequate fitting model, then its deviance D satisfies

$$\frac{D}{\phi} \rightsquigarrow \chi_{n-p}^2$$

To estimate the parameter ϕ , we can use this result to justify the moment-based estimator

$$\hat{\phi} = \frac{D}{n-p}$$

Now, combining these results, we can establish a procedure to perform deviance tests if ϕ is unknown. Consider the statistic

$$\frac{\left[\frac{(D_0 - D_1)}{\phi(p_1 - p_0)} \right]}{\left[\frac{D}{\phi(n-p)} \right]} = \frac{(D_0 - D_1)/(p_1 - p_0)}{D/(n-p)}$$

From above, the RHS is a ratio of two approximately chi-squared random variables divided by their respective degrees of freedom; assuming (with some justification) that $D_0 - D_1$ and D are **independent**, we have by standard results that

$$\frac{(D_0 - D_1)/(p_1 - p_0)}{D/(n-p)} \rightsquigarrow Fisher(p_1 - p_0, n-p)$$

that is, the statistic is approximately Fisher-F distributed. Thus for the analysis of deviance statistic with estimated dispersion

$$\frac{D_0 - D_1}{\hat{\phi}} = (p_1 - p_0) \times \frac{(D_0 - D_1)/(p_1 - p_0)}{D/(n-p)} \rightsquigarrow (p_1 - p_0)F$$

where $F \sim Fisher(p_1 - p_0, n-p)$.

Now, as $\nu \rightarrow \infty$, if $Z_\nu \sim \chi_\nu^2$, then

$$\frac{Z_\nu}{\nu} \xrightarrow{p} 1$$

so therefore we must have that as $F \sim Fisher(\nu_1, \nu_2)$,

$$F \stackrel{d}{=} \frac{Z_{\nu_1}/\nu_1}{Z_{\nu_2}/\nu_2} \xrightarrow{d} Z_{\nu_1}/\nu_1$$

as $\nu_2 \rightarrow \infty$. Therefore, for large n

$$\frac{D_0 - D_1}{\hat{\phi}} \rightsquigarrow \chi_{p_1 - p_0}^2.$$

EXPANSIONS FOR POISSON MODEL

Let $Y \sim \text{Poisson}(\mu)$, and let $g(Y)$ be a differentiable transformation. A Taylor expansion around μ yields

$$g(Y) = g(\mu) + (Y - \mu)\dot{g}(\mu) + \frac{1}{2}(Y - \mu)^2\ddot{g}(\mu) + \text{O}_p((Y - \mu)^3)$$

Truncating at the first order term, as $\mathbb{E}[Y] = \text{Var}[Y] = \mu$,

$$\mathbb{E}[g(Y)] \simeq g(\mu) \quad \text{Var}[g(Y)] = \{\dot{g}(\mu)\}^2\mu$$

Truncating at the second term

$$\mathbb{E}[g(Y)] \simeq g(\mu) + \frac{1}{2}\mathbb{E}[(Y - \mu)^2]\ddot{g}(\mu) = g(\mu) + \frac{\mu}{2}\ddot{g}(\mu)$$

and

$$\text{Var}[g(Y)] \simeq \text{Var}[g(\mu) + (Y - \mu)\dot{g}(\mu) + (Y - \mu)^2\ddot{g}(\mu)/2] = \text{Var}[\dot{g}(\mu)Y + (Y - \mu)^2\ddot{g}(\mu)/2]$$

Now, using the formula for variance $\text{Var}[X] = \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2$, for the first term

$$(\dot{g}(\mu)Y + (Y - \mu)^2\ddot{g}(\mu)/2)^2 = (\dot{g}(\mu)Y + (Y^2 - 2Y + \mu^2)\ddot{g}(\mu)/2)^2 = (a_0 + a_1Y + a_2Y^2)^2$$

where

$$a_0 = \mu^2 \frac{\{\ddot{g}(\mu)\}^2}{4} \quad a_1 = \dot{g}(\mu) - \frac{\{\ddot{g}(\mu)\}^2}{2} \quad a_2 = \frac{\{\ddot{g}(\mu)\}^2}{4}$$

and

$$(a_0 + a_1Y + a_2Y^2)^2 = b_0 + b_1Y + b_2Y^2 + b_3Y^3 + b_4Y^4$$

where

$$b_0 = a_0^2 \quad b_1 = 2a_0a_1 \quad b_2 = 2a_0a_2 + a_1^2 \quad b_3 = 2a_1a_2 \quad b_4 = a_2^2.$$

Thus to compute the $\mathbb{E}[X^2]$ term, we need $\mathbb{E}[Y^r]$ for $r = 1, 2, 3, 4$. We have

$$\begin{aligned} \mathbb{E}[Y] &= \mu \\ \mathbb{E}[Y^2] &= \mu(1 + \mu) \\ \mathbb{E}[Y^3] &= \mu(1 + 3\mu + \mu^2) \\ \mathbb{E}[Y^4] &= \mu(1 + 7\mu + 6\mu^2 + \mu^3) \end{aligned}$$

For the second term, $\{\mathbb{E}[X]\}^2$, we have

$$\mathbb{E}[(\dot{g}(\mu)Y + (Y - \mu)^2\ddot{g}(\mu)/2)] = \dot{g}(\mu)\mu + \frac{\mu\ddot{g}(\mu)}{2}$$

which we need to square.

Square-root transformation: $g(t) = t^{1/2}$

$$\dot{g}(t) = \frac{1}{2}t^{-1/2} \quad \ddot{g}(t) = -\frac{1}{4}t^{-3/2}$$

so therefore

$$a_0 = \frac{1}{32\mu} \quad a_1 = \frac{1}{2\mu^{1/2}} - \frac{1}{16\mu^3} \quad a_2 = \frac{1}{32\mu^3}$$

3.5 STRUCTURED LOG-LINEAR MODELS

EXAMPLE: MODELS FOR SQUARE CONTINGENCY TABLES

The following data represent recorded voting for a cohort of 1651 voters in successive elections; each member of the cohort voted for one of four parties in each election:

		Election 2			
		1	2	3	4
Election 1	1	850	35	25	6
	2	9	286	21	6
	3	3	26	185	5
	4	3	26	27	138

We can consider independence, quasi-independence, symmetry and quasi-symmetry models; perhaps the last two address the most interesting questions in research terms. The following code can be used to fit these models using the `gnm` library in R:

```
1 library(gnm)
2
3 elect.count<-c(850,35,25,6,9,286,21,6,3,26,185,5,3,26,27,138)
4 elect.one<-gl(4,4)
5 elect.two<-gl(4,1,16)
6
7 elections<-data.frame(count=elect.count,Elect1=elect.one,Elect2=elect.two)
8
9 independence<-glm(count~Elect1+Elect2,family=poisson,data=elections)
10 quasi.indep<-glm(count~Elect1+Elect2+Diag(Elect1,Elect2),
11                  family=poisson,data=elections)
12 symmetry<-glm(count~Symm(Elect1,Elect2),family=poisson,data=elections)
13 quasi.symm<-glm(count~Elect1+Elect2+Symm(Elect1,Elect2),family=poisson,
14                 data=elections)
15 comparison1<-anova(independence,symmetry,quasi.indep,quasi.symm)
16 print(comparison1)
17
18 > print(comparison1)
19 Analysis of Deviance Table
20
21 Model 1: count ~ Elect1 + Elect2
22 Model 2: count ~ Elect1 + Elect2 + Diag(Elect1, Elect2)
23 Model 3: count ~ Symm(Elect1, Elect2)
24 Model 4: count ~ Elect1 + Elect2 + Symm(Elect1, Elect2)
25
26 Resid. Df Resid. Dev Df Deviance
27 1          9      2461.65
28 2          5         5.18  4   2456.48
29 3          6        67.81 -1   -62.64
30 4          3         2.49  3    65.32
```

The following results are obtained for the fits of the models (p is the number of parameters fitted):
On deviance terms, clearly the two quasi- models fit better, and each appears to be an adequate model given the deviance/residual degrees of freedom comparison. If a direct comparison were possible, then comparing the quasi-independence model (simpler) with the quasi-symmetry model (more complex) would be based on $\Delta\text{Dev.} = 2.682$, $\Delta\text{df} = 2$, implying a conclusion that we cannot reject the

Model	Deviance	Res. Df.	p
Independence	2461.655	9	7
Quasi-Independence	5.175	5	11
Symmetry	67.814	6	10
Quasi-symmetry	2.494	3	13

hypothesis that the simpler model is an adequate simplification of the more complex model. However, **these models are not nested.**

Recall the definitions of the quasi- models: taking the $j = k = 1$ case as the baseline group, and utilizing the usual identifiability constraints, we have

- **Quasi-Independence:**

$$\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_j \mathbb{1}\{j = k\} = \begin{cases} \beta + \gamma_1 & j = 1, k = 1 \\ \beta + \beta_j^{(R)} & j = 2, 3, 4, k = 1 \\ \beta + \beta_k^{(C)} & j = 1, k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_j^{(C)} + \gamma_j & j = k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_k^{(C)} & j = 2, 3, 4, j \neq k. \end{cases}$$

Thus the eleven parameters are

$$\beta, \beta_2^{(C)}, \beta_3^{(C)}, \beta_4^{(C)}, \beta_2^{(R)}, \beta_3^{(R)}, \beta_4^{(R)}, \gamma_1, \gamma_2, \gamma_3, \gamma_4$$

- **Quasi-Symmetry:** $\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk}$, where $\gamma_{jk} = \gamma_{kj}$ for $j \neq k$. That is,

$$\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk} = \begin{cases} \beta & j = 1, k = 1 \\ \beta + \beta_j^{(R)} + \gamma_{jk} & j = 2, 3, 4, k = 1 \\ \beta + \beta_k^{(C)} + \gamma_{jk} & j = 1, k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk} & j, k = 2, 3, 4 \end{cases}$$

Thus the thirteen parameters are

$$\beta, \beta_2^{(C)}, \beta_3^{(C)}, \beta_4^{(C)}, \beta_2^{(R)}, \beta_3^{(R)}, \beta_4^{(R)}, \gamma_{12}, \gamma_{13}, \gamma_{14}, \gamma_{23}, \gamma_{24}, \gamma_{34}$$

It is not possible to restrict the quasi-symmetry model in such a way as to produce the quasi-independence model. On the basis of these results alone, it is not possible to assess which model is more appropriate.

3.6 ITERATIVE PROPORTIONAL FITTING

A Poisson log-linear model is often used to model data in a multi-way contingency table. For a three-way table, defined by three factors A, B and C , it is assumed that $Y_{jkl} \sim \text{Poisson}(\mu_{jkl})$ where the saturated model can be written

$$\log \mu_{jkl} = \beta_0 + \beta_j^{(A)} + \beta_k^{(B)} + \beta_l^{(C)} + \beta_{jk}^{(AB)} + \beta_{jl}^{(AC)} + \beta_{kl}^{(BC)} + \beta_{jkl}^{(ABC)}$$

and all other log-linear models are nested inside the saturated model. If the factors take J, K and L levels, then there are a maximum of JKL parameters. The log likelihood is

$$l_n(\beta) = \sum_j \sum_k \sum_l (y_{jkl} \log \mu_{jkl} - \mu_{jkl}) + \text{const.}$$

Constraints for identifiability must be introduced: for example, sum-to-zero constraints include

$$\begin{aligned} \text{Main effects:} \quad & \sum_{j=1}^J \beta_j^{(A)} = 0 \\ \text{Interactions:} \quad & \sum_{k=1}^K \beta_{kl}^{(AB)} = 0 \quad \text{for each } l, \\ & \sum_{k=1}^K \sum_{l=1}^L \beta_{jkl}^{(ABC)} = 0 \quad \text{for each } j \end{aligned}$$

etc. To find the ML estimates, note first that for **any** of the parameters β that appear in the form of μ_{jkl} , for the Poisson model, we have

$$\frac{\partial \mu_{jkl}}{\partial \beta} = \mu_{jkl}.$$

Therefore, solving the score equation for β_0 reduces to

$$\frac{\partial l_n(\beta)}{\partial \beta_0} = \sum_j \sum_k \sum_l (y_{jkl} - \mu_{jkl}) = 0 \quad (9)$$

But note that $\mu_{jkl} = \exp\{\beta_0 + (\beta_j^{(A)} + \dots + \beta_{jkl}^{(ABC)})\} = \exp\{\beta_0\} \omega_{jkl}$, say, so equation (9)

$$\exp\{\hat{\beta}_0\} = \frac{\sum_j \sum_k \sum_l y_{jkl}}{\sum_j \sum_k \sum_l \omega_{jkl}} = \frac{n}{\omega_{\dots}}$$

say, where n is the total count, whatever the ω values are. Therefore the log likelihood using this reparameterization can be written, up to an additive constant, as

$$l_n(\omega) = \sum_j \sum_k \sum_l \left(y_{jkl} \log \frac{n \omega_{jkl}}{\omega_{\dots}} \right) - n = \sum_j \sum_k \sum_l \left(y_{jkl} \log \frac{\omega_{jkl}}{\omega_{\dots}} \right) + n \log n - n$$

This is a complicated function of the ω parameters, so differentiating and solving to compute the ML estimates is not possible analytically; the exception to this is the case of the **saturated** model.

For a non-saturated model, the Poisson log likelihood is not amenable to direct maximization, but a “block-based” method can be used. Consider the model written in traditional notation as

$$A + B + C + A.B + A.C + B.C.$$

In this model, we have that

$$\log \mu_{jkl} = \beta_0 + \beta_j^{(A)} + \beta_k^{(B)} + \beta_l^{(C)} + \beta_{jk}^{(AB)} + \beta_{jl}^{(AC)} + \beta_{kl}^{(BC)}$$

and the identifiability constraints are

$$\sum_j \beta_j^{(A)} = 0 \quad \sum_k \beta_{jk}^{(AB)} = 0 \quad \sum_l \beta_{jl}^{(AC)} = 0$$

and similarly for the other factors.

Step 1: We write

$$\mu_{jkl} = \omega_{jk} \alpha_{jkl}$$

where

$$\omega_{jk} = \exp\{\beta_0 + \beta_j^{(A)} + \beta_k^{(B)} + \beta_{jk}^{(AB)}\} \quad \alpha_{jkl} = \exp\{\beta_l^{(C)} + \beta_{jl}^{(AC)} + \beta_{kl}^{(BC)}\}$$

and consider the score equation for a single parameter, $\beta_{jk}^{(AB)}$. We have for each j and k

$$\frac{\partial l_n(\beta)}{\partial \beta_{jk}^{(AB)}} = \sum_l (y_{jkl} - \mu_{jkl}) = y_{jk.} - \mu_{jk.} \quad (10)$$

say.

Step 2: We presume that the α_{jkl} terms are **known**. We have $\mu_{jkl} = \omega_{jk} \alpha_{jkl}$, so therefore solving (10) equal to zero yields $\hat{\omega}_{jk} = y_{jk.} / \alpha_{jk.}$. By similar calculations, we can also conclude that

$$\frac{\partial l_n(\beta)}{\partial \beta_0} = \sum_j \sum_k \left\{ \sum_l (y_{jkl} - \mu_{jkl}) \right\} = 0 \quad \frac{\partial l_n(\beta)}{\partial \beta_j^{(A)}} = \sum_k \left\{ \sum_l (y_{jkl} - \mu_{jkl}) \right\} = 0 \quad \frac{\partial l_n(\beta)}{\partial \beta_k^{(B)}} = \sum_j \left\{ \sum_l (y_{jkl} - \mu_{jkl}) \right\} = 0.$$

These remaining score equations can be solved analytically for known α_{jkl} :

$$\begin{aligned} \hat{\beta}_0 &= \frac{1}{JK} \sum_j \sum_k \log \hat{\omega}_{jk} = \frac{1}{JK} \sum_j \sum_k \log \left\{ \frac{y_{jk.}}{\alpha_{jk.}} \right\} \\ \hat{\beta}_j^{(A)} &= \frac{1}{K} \sum_k \log \hat{\omega}_{jk} - \hat{\beta}_0 \\ \hat{\beta}_k^{(B)} &= \frac{1}{J} \sum_j \log \hat{\omega}_{jk} - \hat{\beta}_0 \\ \hat{\beta}_{jk}^{(AB)} &= \log \hat{\omega}_{jk} - \hat{\beta}_j^{(A)} - \hat{\beta}_k^{(B)} - \hat{\beta}_0 \end{aligned}$$

These parameters are thus computed according to these formulae, and satisfy the identifiability constraints.

Step 3: To proceed, we consider now holding parameters

- $\beta_0, \beta_j^{(A)}, \beta_l^{(C)}, \beta_{jl}^{(AC)}$
- $\beta_0, \beta_k^{(B)}, \beta_l^{(C)}, \beta_{kl}^{(BC)}$

constant in turn, and then compute the parameter updates are recursively until convergence. Note that the parameter blocks here are not disjoint.

It can be shown that this algorithm leads to a non-decreasing sequence of log likelihood values. The algorithm is termed *Iterative Proportional Fitting*.

3.7 OVERDISPERSION AND UNDERDISPERSION

THE NEGATIVE BINOMIAL MODEL

Recall the tumour data, with two factors

- Tumour type: `type`
 - 1- Hutchinsons melanotic freckle, 2- Superficial spreading melanoma, 3- Nodular, 4- Indeterminate
 - Location: `site`
 - 1- Head and Neck, 2- Trunk, 3-Extremities
- defining a 4×3 table.

		Site			Total
		1	2	3	
Tumor	1	22	2	10	34
	2	16	54	115	185
	3	19	33	73	125
	4	11	17	28	56
Total		68	106	226	400

We can compare the parameter estimates for the main effects (independence) model

`type+site`

under the two assumptions

- M_P : Y Poisson distributed
- M_N : Y Negative binomially distributed.

Recall that the negative binomial model can be generated using a Gamma mixture model:

$$Y_{jk}|Z_{jk} = z_{jk} \sim \text{Poisson}(z_{jk}\mu_{jk}) \quad Z_{jk} \sim \text{Gamma}(\theta_Z, \theta_Z)$$

with the Z_{jk} random variables mutually independent, and the Y_{jk} conditionally independent given the Z_{jk} . The marginal pmf of Y is easily computed to be

$$f_Y(y; \theta_Z, \mu) = \frac{\Gamma(\theta_Z + y)}{\Gamma(\theta_Z)y!} \left(\frac{\mu}{\theta_Z + \mu} \right)^y \left(\frac{\theta_Z}{\theta_Z + \mu} \right)^{\theta_Z} \quad y = 0, 1, 2, \dots$$

for parameters $\mu, \theta_Z > 0$. Then $\mathbb{E}[Y] = \mu$, but $\text{Var}[Y] = \mu + \mu^2/\theta_Z$. If θ_Z is large, then the Poisson and negative binomial model become very similar. A log link function $\eta = \log \mu$ can be used for both models.

The following code can be used to fit the models, including the estimation of θ_Z .

```
1 site<-gl(3,1,12);tumor<-gl(4,3)
2 Y<-c(22,2,10,16,54,115,19,33,73,11,17,28)
3 cancer<-data.frame(tumor,site,Y)
4
5 #Poisson fit
6 cancer.mp <- glm(Y ~ tumor + site, family=poisson,data=cancer)
7 #Negative Binomial fit
8 library(MASS)
9 cancer.mn<-glm.nb(Y ~ tumor+site, data=cancer)
```

Results (edited) are as follows:

```

1 Call: glm(formula = Y ~ tumor + site, family = poisson, data = cancer)
2
3 Coefficients:
4             Estimate Std. Error z value Pr(>|z|)
5 (Intercept)   1.7544     0.2040   8.600 < 2e-16 ***
6 tumor2        1.6940     0.1866   9.079 < 2e-16 ***
7 tumor3        1.3020     0.1934   6.731 1.68e-11 ***
8 tumor4        0.4990     0.2174   2.295  0.02173 *
9 site2         0.4439     0.1554   2.857  0.00427 **
10 site3        1.2010     0.1383   8.683 < 2e-16 ***
11 ---
12 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
13
14 (Dispersion parameter for poisson family taken to be 1)
15
16 Null deviance: 295.203  on 11  degrees of freedom
17 Residual deviance: 51.795  on  6  degrees of freedom
18 AIC: 122.91
19
20 Call: glm.nb(formula = Y ~ tumor + site, data = cancer, init.theta =
21 5.020142616, link = log)
22
23 Coefficients:
24             Estimate Std. Error z value Pr(>|z|)
25 (Intercept)   2.20203     0.37444   5.881 4.08e-09 ***
26 tumor2        1.48351     0.40991   3.619 0.000296 ***
27 tumor3        1.12423     0.41330   2.720 0.006526 **
28 tumor4        0.37813     0.42544   0.889 0.374123
29 site2         0.05471     0.35731   0.153 0.878308
30 site3        0.80075     0.34751   2.304 0.021210 *
31 ---
32 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
33
34 (Dispersion parameter for Negative Binomial(5.0201) family taken to be 1)
35
36 Null deviance: 41.328  on 11  degrees of freedom
37 Residual deviance: 13.663  on  6  degrees of freedom
38 AIC: 108.12
39
40             Theta:  5.02
41             Std. Err.:  2.75
42
43 2 x log-likelihood:  -94.124

```

The `glm.nb` call (line 20) fits the negative binomial model using an iterative procedure. The resulting estimate (standard error) of θ_Z is 5.02 (2.75) - see lines 40-41; this is not too large, indicating that the negative binomial model is justified over the Poisson model. Note that the parameter estimates and standard errors change, as do the fitted values (below).

The difference between the $Poisson(\mu)$, and $NegativeBinomial(\mu, \theta_Z)$ is shown below in the figure, with $\mu = \exp\{2\}$ (top panel) and $\mu = \exp\{4\}$ (bottom panel) for $\theta_Z = 5$.

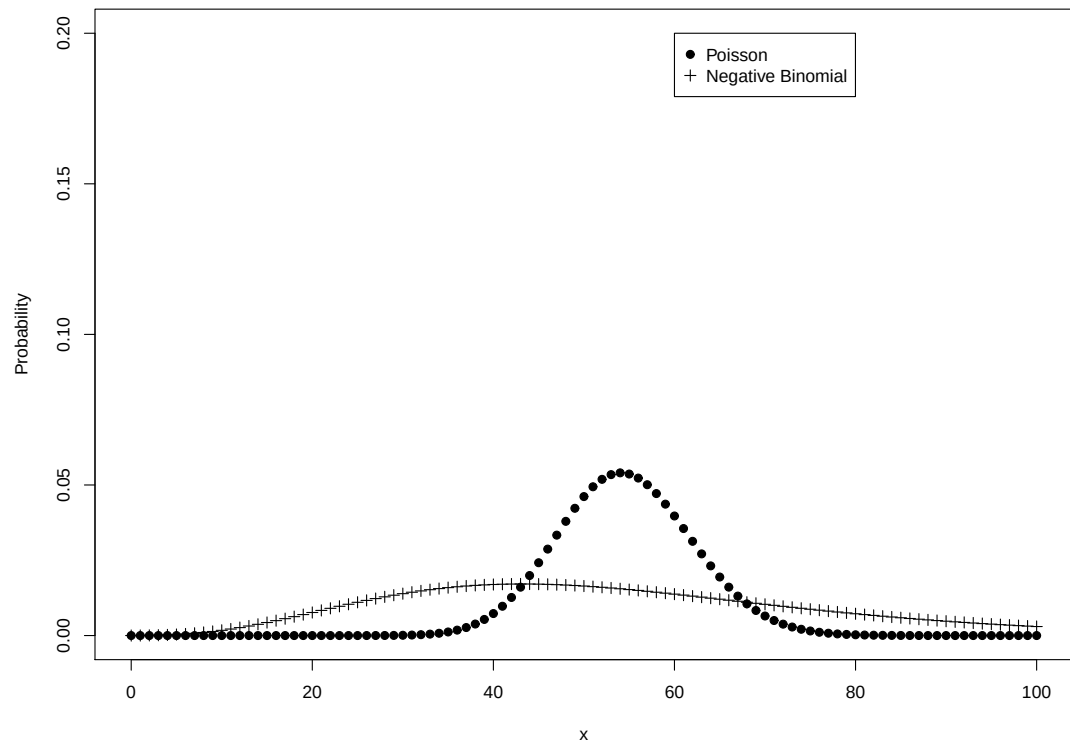
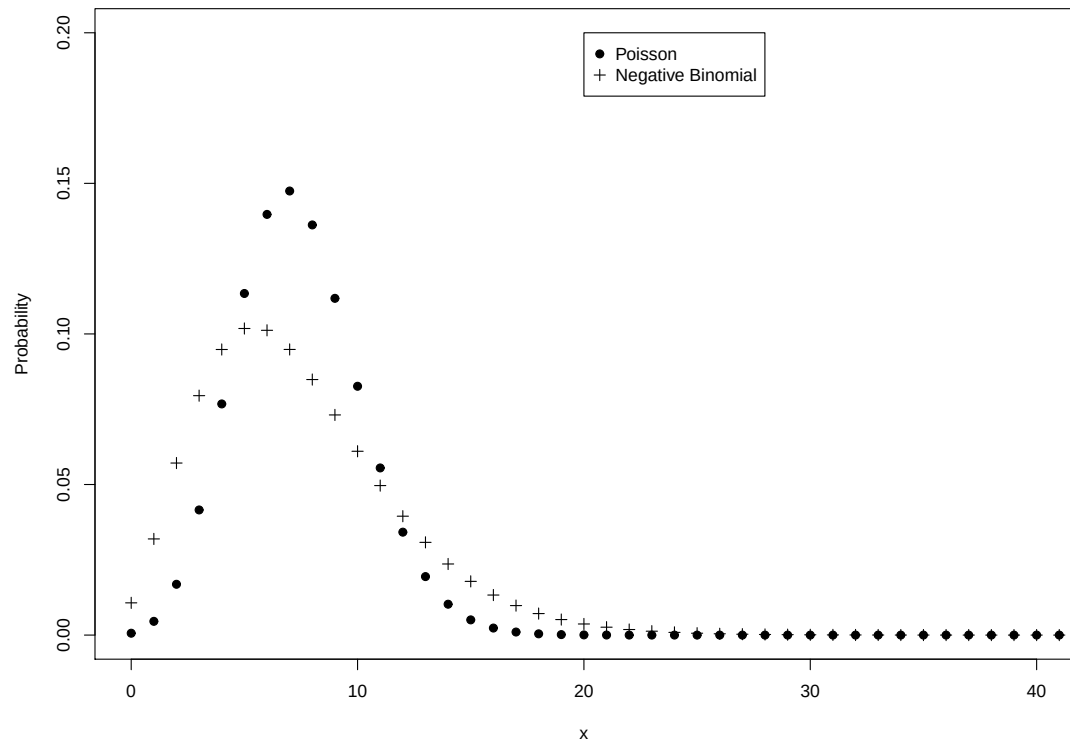


Figure 5: Comparison of Poisson and negative binomial models for different values of μ : $\mu = \exp\{2\}$ (top) and $\mu = \exp\{4\}$ (bottom) for $\theta_Z = 5$.

		Poisson Fit			Neg. Binomial Fit		
		1	2	3	1	2	3
Tumor	1	5.78	9.01	19.21	9.04	9.55	20.14
	2	31.45	49.02	104.52	39.87	42.11	88.79
	3	21.25	33.12	70.62	27.83	29.40	61.99
	4	9.52	14.84	31.64	13.20	13.94	29.40

Table 1: Table of Fitted Values

Model comparison for the Poisson and Negative Binomial models **can** be carried out using a likelihood ratio test as the Poisson model is nested inside the Negative Binomial: if $\varphi = 1/\theta_Z$, then we have that $\varphi = 0$ yields the Poisson, so the likelihood ratio test is

$$H_0 : \varphi = 0$$

$$H_1 : \varphi > 0$$

Therefore, the likelihood ratio statistic is

$$-2 \log \lambda(\mathbf{y}) = -2(l_n^{(P)}(\hat{\mu}) - l_n^{(N)}(\hat{\mu}, \hat{\varphi}))$$

where $l_n^{(P)}$ and $l_n^{(N)}$ are the log likelihoods for Poisson and binomial models.

This model, however, is **non-regular**: the nesting structure is obtained by setting φ equal to a value on the boundary of the parameter space, and hence the regularity assumptions required for the usual asymptotic behaviour are not met. It can be shown that in this case

$$-2 \log \lambda(\mathbf{y}) \xrightarrow{d} V$$

where V is a random variable with a **mixture density**

$$V \stackrel{d}{=} \frac{1}{2} \delta_{\{v=0\}} + \frac{1}{2} \chi_1^2$$

that is with probability mass 1/2 at $v = 0$, and 1/2 on the χ_1^2 density. Under this distribution

$$\Pr[V > v] = \frac{1}{2} S_1(v)$$

where $S_1(v)$ is the tail probability of exceeding v for a χ_1^2 density. $\Pr[V > v]$ is the p -value in the likelihood ratio test.

For these data, we can obtain the optimized log-likelihood values via the `logLik` function in R:

```
> logLik(cancer.mp); logLik(cancer.mn)
'log Lik.' -55.45321 (df=6)
'log Lik.' -47.06196 (df=7)
```

have $v = 16.78249$. Therefore the probability of observing $V > v$ is

```
> 0.5*(1-pchisq(16.78249,1))
[1] 2.095916e-05
```

hence a highly significant result. This implies that the Negative Binomial Model fits significantly better than the Poisson.

4 MODELS FOR BINOMIAL AND MULTINOMIAL DATA

4.1 BINOMIAL AND LOGISTIC REGRESSION

LOGISTIC REGRESSION WITH A SINGLE COVARIATE

EXAMPLE: Crabs Data:- 173 observations of the shell widths of female horseshoe crabs, with response value being the number of male “satellite” crabs discovered within a given distance.

- Response Y : Indicator of number of satellites being greater than zero.
- Covariate x : Shell width

These data are available from

`www.math.mcgill.ca/dstephens/523/Data/Crabs.txt`

The following code performs the analysis of two competing models: let

$$\pi(x) = \Pr[Y = 1|x]$$

- m1: Linear

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x$$

- m2: Quadratic

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x + \beta_2 x^2$$

The code also produces a plot using categorized data: the deciles of width are used to define ten categories, and the proportion of crabs that have response 0 and 1 in each category is recorded. This proportion is then plotted against the mean width for the category.

```
1 Crabs<-read.table("Crabs.txt",header=T)
2
3 m1<-glm(I(satell>0)~width,data=Crabs,family=binomial)
4 m2<-glm(I(satell>0)~width+I(width^2),data=Crabs,family=binomial)
5
6 ncat<-10
7 wcat<-cut(Crabs$width,quantile(Crabs$width,prob=c(0:ncat)/ncat),include.lowest=T)
8 wi<-as.numeric(wcat)
9 wm<-rep(0,ncat)
10 for(k in 1:ncat){
11     wm[k]<-mean(Crabs$width[wi == k])
12 }
13 wtab<-table(wcat,I(Crabs$satell>0))
14 prop<-wtab/apply(wtab,1,sum)
15
16 plot(wm,prop[,2],ylim=range(0,1),
17 xlab="x=Width",ylab=expression(pi(x)),pch=19,xlim=range(22,32))
18 xv<-seq(22,32,by=0.01)
19 lp1<-m1$coef[1]+xv*m1$coef[2]
20 p1<-exp(lp1)/(1+exp(lp1))
21 lines(xv,p1)
22 lp2<-m2$coef[1]+xv*m2$coef[2]+xv^2*m2$coef[3]
23 p2<-exp(lp2)/(1+exp(lp2))
24 lines(xv,p2,lty=2)
25 legend(28,0.4,c("Linear","Quadratic"),lty=c(1,2))
```

The categorized data are as follows

Category	Y	
	0	1
[21.0,23.7]	14	5
(23.7,24.5]	10	8
(24.5,25.1]	5	10
(25.1,25.7]	10	9
(25.7,26.1]	5	11
(26.1,26.7]	7	11
(26.7,27.4]	4	12
(27.4,28.2]	4	12
(28.2,29.0]	3	19
(29.0,33.5]	0	14

Results:

```
> summary(m1)
```

```
Call: glm(formula = I(satell > 0) ~ width, family = binomial, data = Crabs)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0281	-1.0458	0.5480	0.9066	1.6942

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-12.3508	2.6287	-4.698	2.62e-06 ***
width	0.4972	0.1017	4.887	1.02e-06 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 225.76 on 172 degrees of freedom
 Residual deviance: 194.45 on 171 degrees of freedom
 AIC: 198.45

```
> summary(m2)
```

```
Call: glm(formula = I(satell > 0) ~ width + I(width^2), family = binomial,
  data = Crabs)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1185	-1.0441	0.5067	0.9483	1.5406

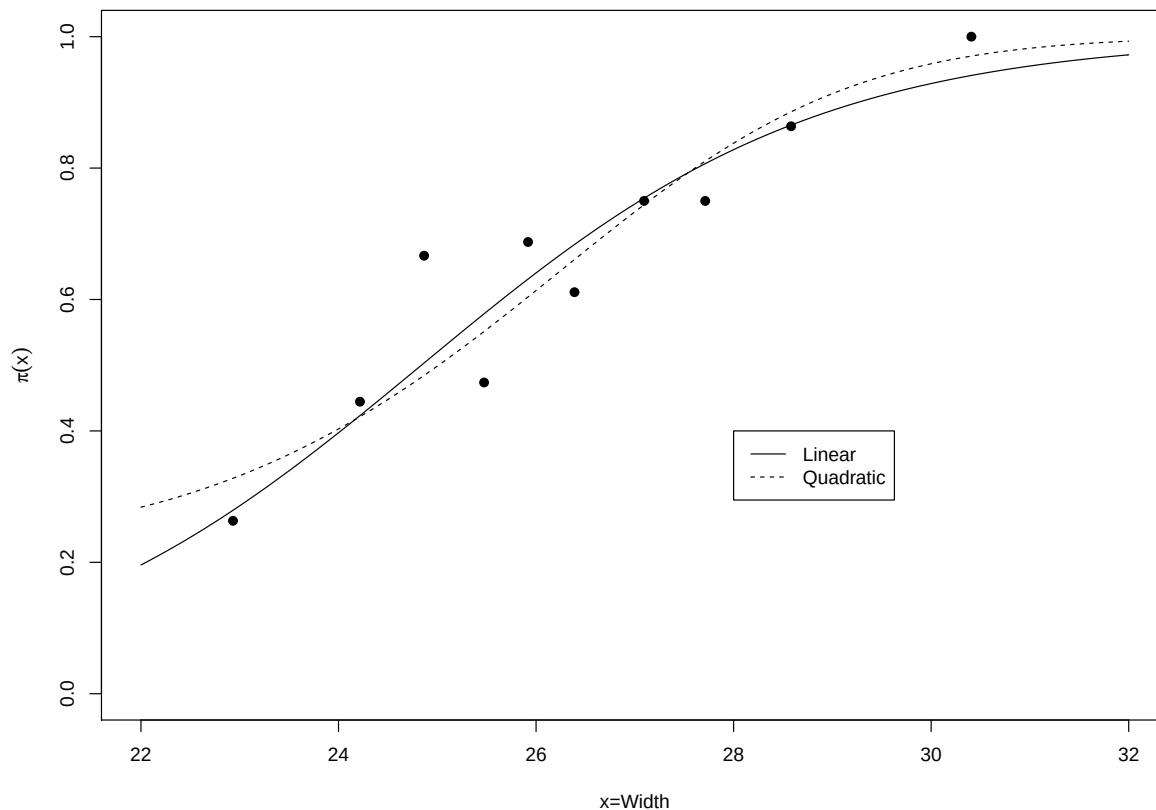
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	14.59156	30.22371	0.483	0.629
width	-1.59572	2.35195	-0.678	0.497
I(width^2)	0.04047	0.04566	0.886	0.375

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 225.76 on 172 degrees of freedom
 Residual deviance: 193.63 on 170 degrees of freedom
 AIC: 199.63

Number of Fisher Scoring iterations: 5



Comparing the two models via their deviances, or by inspecting the coefficient of the quadratic term in model m2, it seems that the linear model does not provide an inferior fit compared to the quadratic model.

```
> anova(m1,m2)
Analysis of Deviance Table
```

```
Model 1: I(satell > 0) ~ width
Model 2: I(satell > 0) ~ width + I(width^2)
  Resid. Df Resid. Dev Df Deviance
1       171      194.45
2       170      193.63  1   0.82542
```

Recall, however, that comparing models via their deviance must be done with care for binary response data. Therefore the more reliable test is based on the Wald statistic

$$W_n = \frac{\hat{\beta}_2}{\text{s.e.}(\hat{\beta}_2)}$$

where $W_n^2 \rightsquigarrow \chi_1^2$ for large n .

ANALYSIS BASED ON FACTOR PREDICTORS

EXAMPLE: AIDS and AZT:- The following data are from a study on the effects of drug AZT in slowing the development of AIDS symptoms. In the study, 338 veterans with compromised immune systems after infection with the AIDS virus were randomly assigned either to receive AZT immediately, or to wait until their T cells showed severe immune weakness. The table below cross-classifies the veterans race, whether they received AZT immediately, and whether they developed AIDS symptoms during the 3-year study.

- Response Y : Indicator of number of developing AIDS symptoms.
- Factor `azt`: Recipient of AZT therapy (Yes, No)
- Factor `race`: Race of subject (White, Black)

Race	AZT	Symptoms	
		Yes	No
White	Yes	14	93
	No	32	81
Black	Yes	11	52
	No	12	43

We have a $2 \times 2 \times 2$ table of counts; via the randomization, the sampling design effectively fixes the number of subjects receiving/not receiving AZT (we have $107+63=170$ receiving AZT, $113+55=168$ not receiving AZT). However, we can fit logistic regression models to the individual response data, taking `azt` and `race` as factor predictors taking two levels.

We may fit five models to the binary response data

```
m0 Null
m1 azt
m2 race
m3 azt+race
m4 azt*race = azt+race+azt:race
```

Code to do this in R is as follows:

```
1 AZT.counts<-c(14, 93, 32, 81, 11, 52, 12, 43)
2 AZT.acounts<-c(sum(AZT.counts[1:2]), sum(AZT.counts[3:4]),
3               sum(AZT.counts[5:6]), sum(AZT.counts[7:8]))
4 AZT.aids<-as.factor(rep(c("Y", "N", "Y", "N", "Y", "N", "Y", "N"), AZT.counts))
5 AZT.azt<-as.factor(rep(c("Y", "N", "Y", "N"), AZT.acounts))
6 AZT.race<-as.factor(rep(c("W", "B"), c(sum(AZT.counts[1:4]), sum(AZT.counts[5:8]))))
7
8 table(AZT.azt, AZT.aids, AZT.race)
9
10 AZT<-data.frame(AZT.race, AZT.azt, AZT.aids)
11 names(AZT)<-c("race", "azt", "aids")
12
13 m0<-glm(aids~1, data=AZT, family=binomial)
14 m1<-glm(aids~azt, data=AZT, family=binomial)
15 m2<-glm(aids~race, data=AZT, family=binomial)
16 m3<-glm(aids~azt+race, data=AZT, family=binomial)
17 m4<-glm(aids~azt*race, data=AZT, family=binomial)
```

Results:

```
> anova(m0,m1,m2,m3,m4)
Analysis of Deviance Table
```

```
Model 1: aids ~ 1
Model 2: aids ~ azt
Model 3: aids ~ race
Model 4: aids ~ azt + race
Model 5: aids ~ azt * race
  Resid. Df Resid. Dev Df Deviance
1      337      342.12
2      336      335.19  1    6.9293
3      336      342.02  0   -6.8338
4      335      335.15  1    6.8709
5      334      333.77  1    1.3835
```

The immediate conclusion of this analysis is that the interaction term `azt.race` can be omitted. The results for the additive model `azt+race` are

```
> summary(m3)
```

```
Call: glm(formula = aids ~ azt + race, family = binomial, data = AZT)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7854	-0.7668	-0.5694	-0.5549	1.9733

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.07357	0.26294	-4.083	4.45e-05	***
aztY	-0.71946	0.27898	-2.579	0.00991	**
raceW	0.05548	0.28861	0.192	0.84755	

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 342.12 on 337 degrees of freedom
Residual deviance: 335.15 on 335 degrees of freedom
AIC: 341.15

Number of Fisher Scoring iterations: 4

This indicates a strong effect of AZT therapy (significant coefficient for `aztY`, estimate (s.e.) is -0.71946 (0.27898)). Comparing the models `m1` and `m3` yields an assessment of whether there is a need for the factor `race` to be included:

```
> anova(m1,m3)
Analysis of Deviance Table
```

```
Model 1: aids ~ azt
Model 2: aids ~ azt + race
  Resid. Df Resid. Dev Df Deviance
1      336      335.19
2      335      335.15  1  0.037084
```

which, in conjunction with the magnitude of the coefficient of `raceW` in the `m3` analysis leads us to conclude that `race` does not need to be included in the model.

Finally, the models `m1` yields an assessment of the magnitude of the AZT effect ignoring `race`, that is, pooling the two race groups.

```
> summary(m1)

Call:glm(formula = aids ~ azt, family = binomial, data = AZT)

Deviance Residuals:
    Min       1Q   Median       3Q      Max 
-0.7793 -0.7793 -0.5640 -0.5640  1.9580 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.0361     0.1755  -5.904 3.54e-09 ***
aztY          -0.7218     0.2787  -2.590 0.00961 **
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 342.12  on 337  degrees of freedom
Residual deviance: 335.19  on 336  degrees of freedom
AIC: 339.19

Number of Fisher Scoring iterations: 4
```

Interpretation: The additive main effects model `azt+race` models the log odds of developing AIDS symptoms as follows: coding `aids=Y` as 1, `aids=N` as 0, the models seek to explain

$$\log\left(\frac{\pi_{1jk}}{\pi_{0jk}}\right) = \log\left(\frac{\pi_{1jk}}{1 - \pi_{1jk}}\right)$$

and we have

$$\log\left(\frac{\pi_{1jk}}{\pi_{0jk}}\right) = \beta + \beta_j^{(azt)} + \beta_k^{(race)} = \begin{cases} \beta & j = 1, k = 1 \quad (azt=N, race=black) \\ \beta + \beta_2^{(azt)} & j = 2, k = 1 \quad (azt=Y, race=black) \\ \beta + \beta_2^{(race)} & j = 1, k = 2 \quad (azt=N, race=white) \\ \beta + \beta_2^{(azt)} + \beta_2^{(race)} & j = 2, k = 2 \quad (azt=Y, race=white) \end{cases}$$

with estimates (from the `summary(m3)` table from the previous page),

$$\hat{\beta} = -1.07357 \quad \hat{\beta}_2^{(azt)} = -0.71946 \quad \hat{\beta}_2^{(race)} = 0.05548$$

Rearranging the formulae above, we have that

$$\log\left(\frac{\pi_{121}/\pi_{021}}{\pi_{111}/\pi_{011}}\right) = \beta_2^{(azt)} \quad \log\left(\frac{\pi_{122}/\pi_{022}}{\pi_{111}/\pi_{011}}\right) = \beta_2^{(azt)} + \beta_2^{(race)} \quad \log\left(\frac{\pi_{122}/\pi_{022}}{\pi_{121}/\pi_{021}}\right) = \beta_2^{(race)}$$

Thus, we conclude that as the coefficient $\beta_2^{(race)}$ has estimate (standard error) 0.05548 (0.28861), with z value 0.192 and p -value 0.84755, there is no evidence to support that the odds-ratio comparing the `azt=Y` and `azt=N` levels differs across the races.

EXAMPLE: Low Birthweight data:- The `birthwt` data frame in R has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Mass during 1986. This data frame contains the following columns:

- `low` indicator of birth weight less than 2.5 kg (binary response).
- `age` mother's age in years (covariate).
- `lwt` mother's weight in pounds at last menstrual period (covariate).
- `race` mother's race (1 = white, 2 = black, 3 = other) (factor with three levels).
- `smoke` smoking status during pregnancy (factor with two levels).
- `ptl` number of previous premature labours.
- `ht` history of hypertension (factor with two levels).
- `ui` presence of uterine irritability (factor with two levels).
- `ftv` number of physician visits during the first trimester.
- `bwt` birth weight in grams (continuous version of response).

These data are available in the MASS library in R. Following the recoding and omission of variables devised by Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S, Fourth ed.*, the following code sets up and performs the analysis:

```
1 attach(birthwt)
2 race <- factor(race, labels = c("white", "black", "other"))
3 ptd <- factor(ptl > 0)
4 ftv <- factor(ftv)
5 levels(ftv)[-1:2] <- "2+"
6 bwt <- data.frame(low = factor(low), age, lwt, race,
7   smoke = (smoke > 0), ptd, ht = (ht > 0), ui = (ui > 0), ftv)
8 detach("birthwt")
9 lbw<-glm(low ~ ., family=binomial, bwt)
```

The model fitted is an additive model; the results are as follows:

```
> summary(lbw)
Call: glm(formula = low ~ ., family = binomial, data = bwt)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.82302	1.24471	0.661	0.50848
age	-0.03723	0.03870	-0.962	0.33602
lwt	-0.01565	0.00708	-2.211	0.02705 *
raceblack	1.19241	0.53597	2.225	0.02609 *
raceother	0.74069	0.46174	1.604	0.10869
smokeTRUE	0.75553	0.42502	1.778	0.07546 .
ptdTRUE	1.34376	0.48062	2.796	0.00518 **
htTRUE	1.91317	0.72074	2.654	0.00794 **
uiTRUE	0.68019	0.46434	1.465	0.14296
ftv1	-0.43638	0.47939	-0.910	0.36268
ftv2+	0.17901	0.45638	0.392	0.69488

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 234.67 on 188 degrees of freedom

Residual deviance: 195.48 on 178 degrees of freedom

AIC: 217.48

We can attempt to omit individual terms by using the `drop1` command

```
> drop1(lbw, test="Chisq")
Single term deletions
```

Model:

```
low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		195.48	217.48		
age	1	196.42	216.42	0.9419	0.331796
lwt	1	200.95	220.95	5.4739	0.019302 *
race	2	201.23	219.23	5.7513	0.056380 .
smoke	1	198.67	218.67	3.1982	0.073717 .
ptd	1	203.58	223.58	8.1085	0.004406 **
ht	1	202.93	222.93	7.4584	0.006314 **
ui	1	197.59	217.59	2.1100	0.146342
ftv	2	196.83	214.83	1.3582	0.507077

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

On this basis, we can attempt to omit several terms, specifically ftv, age and ui.

```
> lbw.1 <- update(lbw, ~. - ftv - age - ui)
> anova(lbw.1, lbw)
```

Analysis of Deviance Table

Model 1: low ~ lwt + race + smoke + ptd + ht

Model 2: low ~ age + lwt + race + smoke + ptd + ht + ui + ftv

	Resid. Df	Resid. Dev	Df	Deviance
1	182	200.48		
2	178	195.48	4	5.0067

```
> summary(lbw.1)
```

```
Call: glm(formula = low ~ lwt + race + smoke + ptd + ht, family = binomial,
data = bwt)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.09462	0.95704	0.099	0.92124
lwt	-0.01673	0.00695	-2.407	0.01608 *
raceblack	1.26372	0.52933	2.387	0.01697 *
raceother	0.86418	0.43509	1.986	0.04701 *
smokeTRUE	0.87611	0.40071	2.186	0.02879 *
ptdTRUE	1.23144	0.44625	2.760	0.00579 **
htTRUE	1.76744	0.70841	2.495	0.01260 *

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 234.67 on 188 degrees of freedom

Residual deviance: 200.48 on 182 degrees of freedom

AIC: 214.48

The deviance change of 5.0067 on 4 degrees of freedom is not a significant result, thus the simpler model with the variables omitted is an adequate simplification. It is possible to investigate further by including interactions between factor predictors, and polynomial terms in the covariate lwt.

4.2 CONNECTION BETWEEN POISSON AND MULTINOMIAL LIKELIHOODS

Binomial: If $Y_1 \sim \text{Poisson}(\mu_1)$ and $Y_2 \sim \text{Poisson}(\mu_2)$ are independent, and if $M = Y_1 + Y_2$, then we have that

$$Y_1 | M = m \sim \text{Binomial}(m, \pi)$$

where $\pi = \mu_1 / (\mu_1 + \mu_2)$. Suppose we have a GLM for the two data, where $\log \mu_1 = \beta_0$, $\log \mu_2 = \beta_0 + \beta_1$. The likelihood based on data (y_1, y_2) is given by

$$L(\beta_0, \beta_1; y_1, y_2) = f_{Y_1, Y_2}(y_1, y_2; \beta_0, \beta_1)$$

so that the log-likelihood is

$$\begin{aligned} l(\beta_0, \beta_1; y_1, y_2) = \log L(\beta_0, \beta_1; y_1, y_2) &= y_1 \log \mu_1 - \mu_1 + y_2 \log \mu_2 - \mu_2 + \text{const.} \\ &= y_1 \beta_0 - e^{\beta_0} + y_2(\beta_0 + \beta_1) - e^{\beta_0 + \beta_1} + \text{const.} \\ &= (y_1 + y_2)\beta_0 + y_2\beta_1 - e^{\beta_0}(1 + e^{\beta_1}) + \text{const.} \end{aligned}$$

Reparameterizing to use $\psi = \mu_1 + \mu_2 = e^{\beta_0} + e^{\beta_0 + \beta_1} = e^{\beta_0}(1 + e^{\beta_1})$, we may re-write the likelihood for (ψ, β_1) as

$$l(\psi, \beta_1; y_1, y_2) = (y_1 + y_2) \log \psi - \psi + y_2\beta_1 - (y_1 + y_2) \log(1 + e^{\beta_1}) + \text{const.}$$

and, equivalently, if $m = y_1 + y_2$,

$$\begin{aligned} l(\psi, \beta_1; m, y_2) &= m \log \psi - \psi + y_2\beta_1 - m \log(1 + e^{\beta_1}) + \text{const.} \\ &= l_M(\psi; m) + l_{Y_2|M}(\beta_1; y_2, m) + \text{const.} \end{aligned}$$

where

$$l_M(\psi; m) = m \log \psi - \psi \quad l_{Y_2|M}(\beta_1; y_2, m) = y_2\beta_1 - m \log(1 + e^{\beta_1})$$

The first term $l_M(\psi; m)$ is (up to an additive constant) the log-likelihood for a Poisson model with mean $\psi = e^{\beta_0}(1 + e^{\beta_1}) = \mu_1 + \mu_2$. The second term is (up to an additive constant) the log-likelihood for a Binomial model with parameters m and

$$\pi = \frac{e^{\beta_1}}{1 + e^{\beta_1}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0} + e^{\beta_0 + \beta_1}} = \frac{\mu_2}{\mu_1 + \mu_2}.$$

Hence inference about β_1 can be made via the **conditional** likelihood for Y_2 , given $M = Y_1 + Y_2$ is observed to be m ; this follows as, in the new parameterization, terms involving β_1 appear only in the conditional likelihood.

Multinomial I: This result may be extended to the K random variable case, and a multinomial model. Suppose that $Y_k \sim \text{Poisson}(\mu_k)$ for $k = 1, \dots, K$ are independent, and

$$\log \mu_k = \beta_0 + \beta_1 x_k$$

say (that is, retaining a two-parameter model for the moment). The calculation above goes through as before with

$$\psi = \sum_{k=1}^K e^{\beta_0 + \beta_1 x_k} = e^{\beta_0} \sum_{k=1}^K e^{\beta_1 x_k}$$

and $M = Y_1 + \dots + Y_K$. Then if $m = y_1 + \dots + y_K$,

$$\begin{aligned} l(\psi, \beta_1; m, y_1, \dots, y_{K-1}) &= m \log \psi - \psi + \sum_{k=1}^K y_k \beta_1 x_k - m \log \left(\sum_{k=1}^K e^{\beta_1 x_k} \right) + \text{const.} \\ &= l_M(\psi; m) + l_{Y|M}(\beta_1; y_1, \dots, y_K, m) + \text{const.} \end{aligned}$$

The first term $l_M(\psi; m)$ is (up to an additive constant) the log-likelihood for a Poisson model with mean $\psi = \mu_1 + \dots + \mu_K$. The second term is (up to an additive constant) the log-likelihood for a Multinomial model with parameters m and $(\pi_1, \dots, \pi_K)$ where

$$\pi_k = \frac{e^{\beta_1 x_k}}{\sum_{l=1}^K e^{\beta_1 x_l}} = \frac{e^{\beta_0 + \beta_1 x_k}}{\sum_{l=1}^K e^{\beta_0 + \beta_1 x_l}} = \frac{\mu_k}{\mu_1 + \dots + \mu_K}.$$

We can see this result as for the multinomial case

$$L(\pi_1, \dots, \pi_K; m, y_1, \dots, y_K) = \frac{m!}{y_1! \dots y_K!} \pi_1^{y_1} \dots \pi_K^{y_K}$$

yielding a log-likelihood

$$\begin{aligned} l(\pi_1, \dots, \pi_K; m, y_1, \dots, y_K) &= \sum_{k=1}^K y_k \log \pi_k + \text{const.} = \sum_{k=1}^K y_k \log \left(\frac{e^{\beta_1 x_k}}{\sum_{l=1}^K e^{\beta_1 x_l}} \right) + \text{const.} \\ &= \sum_{k=1}^K y_k \left(\beta_1 x_k - \log \left(\sum_{l=1}^K e^{\beta_1 x_l} \right) \right) + \text{const.} \\ &= \beta_1 \sum_{k=1}^K y_k x_k - m \log \left(\sum_{l=1}^K e^{\beta_1 x_l} \right) + \text{const.} \end{aligned}$$

as required. In the above calculations we may write $\beta_k = \beta_1 x_k$ for convenience.

Multinomial II: Finally, suppose that we have a Poisson model where $\log \mu_k = \beta_0 + \beta_k$, with $\beta_1 = 0$ for identifiability. Writing

$$\psi = \sum_{k=1}^K \mu_k = \sum_{k=1}^K e^{\beta_0 + \beta_k} = e^{\beta_0} \left(1 + \sum_{k=2}^K e^{\beta_k} \right)$$

and conditioning on $M = Y_1 + \dots + Y_K$ taking the value m , following through the identical steps, we have the same factorization

$$l(\psi, \beta_2, \dots, \beta_K; m, y_1, \dots, y_{K-1}) = l_M(\psi; m) + l_{Y|M}(\beta_2, \dots, \beta_K; y_1, \dots, y_K, m) + \text{const.}$$

where

$$\begin{aligned} l_M(\psi; m) &= m \log \psi - \psi \\ l_{Y|M}(\beta_2, \dots, \beta_K; y_1, \dots, y_K, m) &= \sum_{k=1}^K y_k \beta_k - m \log \left(\sum_{l=1}^K e^{\beta_l} \right). \end{aligned}$$

The conditional distribution that yields the log-likelihood $l_{Y|M}(\beta_2, \dots, \beta_K; y_1, \dots, y_K, m)$ is a multinomial distribution with probabilities

$$\pi_k = \frac{e^{\beta_k}}{\sum_{l=1}^K e^{\beta_l}} = \frac{e^{\beta_0 + \beta_k}}{\sum_{l=1}^K e^{\beta_0 + \beta_l}} = \frac{\mu_k}{\mu_1 + \dots + \mu_K}.$$

Note that as $\beta_1 = 0$,

$$\pi_1 = \frac{1}{1 + \sum_{l=2}^K e^{\beta_l}} \quad \pi_k = \frac{e^{\beta_k}}{1 + \sum_{l=2}^K e^{\beta_l}}, \quad k = 2, \dots, K.$$

4.3 2×2 TABLES AND CASE-CONTROL DESIGNS

MEASURES OF EFFECT FROM 2×2 TABLES

For binary response data $Y = 0, 1$, with a single factor predictor $X = 0, 1$ representing exposure or treatment status ($X = 1$ implying treated or exposed, say), a 2×2 table can be constructed for a sample of size n . Depending on the study design, we might assume binomial sampling in the columns (*prospective*) or rows (*retrospective*).

		X		
		0	1	Total
Y	0	y_{00}	y_{01}	$n_{0.}$
	1	y_{10}	y_{11}	$n_{1.}$
Total		$m_{.0}$	$m_{.1}$	n

Possible measure of effect that may be of interest in different contexts include:

- *Rate of Incidence* or *Risk* in Exposed and Unexposed Groups

$$\pi_1 = \Pr[Y = 1|X = 1] \quad \pi_0 = \Pr[Y = 1|X = 0]$$

- *Odds on Incidence* in Exposed and Unexposed Groups

$$\omega_1 = \frac{\pi_1}{1 - \pi_1} \quad \omega_0 = \frac{\pi_0}{1 - \pi_0}$$

- *Risk Difference*

$$RD = \pi_1 - \pi_0$$

- *Risk Ratio* or *Relative Risk*

$$RR = \frac{\pi_1}{\pi_0}$$

- *Odds ratio* (OR)

$$\psi = \frac{\pi_1/\pi_0}{(1 - \pi_1)/(1 - \pi_0)} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} = \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1)}$$

- *Excess Relative Risk*

$$ERR = \frac{\pi_1 - \pi_0}{\pi_0}$$

- *Attributable Risk*

$$AR = \frac{\pi_1 - \pi_0}{\pi_1}$$

All of these quantities are estimable from *prospective* studies, as the parameters π_0, π_1 are estimable; we have that

$$\hat{\pi}_0 = \frac{y_{10}}{m_{.0}} \quad \hat{\pi}_1 = \frac{y_{11}}{m_{.1}}$$

Retrospective (case-control) studies, however, are only able to give information about

$$\varpi_1 = \Pr[X = 1|Y = 1] \quad \varpi_0 = \Pr[X = 1|Y = 0]$$

and ψ , and functions of these parameters.

Under the *rare disease/exposure hypothesis*, however, where the *prevalence* $\phi = \Pr[Y = 1]$, and hence π_0, π_1 are **small**, case-control studies can be used to learn (approximately) about the risk ratio/relative risk π_1/π_0 ; this follows as, in that case, for $j = 0, 1$,

$$\frac{\pi_j}{1 - \pi_j} = \pi_j (1 - \pi_j)^{-1} = \pi_j + \pi_j^2 + \pi_j^3 + \dots \approx \pi_j$$

so that

$$\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \approx \frac{\pi_1}{\pi_0}.$$

By invariance properties of ML estimates, and the *Delta Method*, we can obtain approximate sampling distributions of these effect measures. We have for $j = 0, 1$ that

$$\hat{\pi}_j = \frac{Y_{1j}}{m_{.j}} \doteq \mathcal{N}(\pi_j, \pi_j(1 - \pi_j)/m_{.j})$$

The estimated standard error of $\hat{\pi}_j$ is therefore

$$\sqrt{\frac{\hat{\pi}_j(1 - \hat{\pi}_j)}{m_{.j}}}$$

The Delta method says that for statistic T_n , if

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$$

then if function $g(t)$ has (non-zero) derivative $\dot{g}(t)$

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} Z \sim \mathcal{N}(0, \{\dot{g}(\theta)\}^2 \sigma^2)$$

Using the Delta method with $g(t) = \log t$, where $\dot{g}(t) = 1/t$, we have that the estimated variance of $\log \hat{\pi}_j$ is

$$\frac{(1 - \hat{\pi}_j)}{\hat{\pi}_j m_{.j}} = \frac{m_{.j} - y_{1j}}{m_{.j} y_{1j}} = \frac{1}{y_{1j}} - \frac{1}{m_{.j}}$$

Therefore the estimated variance of the estimator of the log relative risk, $\log(\hat{\pi}_1/\hat{\pi}_0)$ is

$$\left(\frac{1}{y_{10}} - \frac{1}{m_{.0}} \right) + \left(\frac{1}{y_{11}} - \frac{1}{m_{.1}} \right)$$

Similarly, with $g(t) = \log(t/(1 - t))$, so that

$$\dot{g}(t) = \frac{1}{t(1 - t)}$$

we have that the estimated variance of $\log(\hat{\pi}_j/(1 - \hat{\pi}_j))$ is

$$\frac{1}{\hat{\pi}_j(1 - \hat{\pi}_j)m_{.j}} = \frac{m_{.j}}{y_{0j}y_{1j}} = \frac{1}{y_{0j}} + \frac{1}{y_{1j}}$$

as $m_{.j} = y_{0j} + y_{1j}$. Hence the estimated variance of the estimator of the log odds-ratio

$$\log \left(\frac{\hat{\pi}_1/(1 - \hat{\pi}_1)}{\hat{\pi}_0/(1 - \hat{\pi}_0)} \right)$$

is

$$\frac{1}{y_{00}} + \frac{1}{y_{10}} + \frac{1}{y_{01}} + \frac{1}{y_{11}}$$

EXAMPLES - MULTINOMIAL DATA

EXAMPLE: Birth malformation data:- The following table contains the counts of three different birth outcomes (in terms of level of malformation) in an animal study of drug toxicity

Dose Level	Outcome		
	1	2	3
0	15	1	281
62.5	17	0	225
125	22	7	283
250	38	59	202
500	144	132	9

One way to fit these data is to regard it as a log-linear model;

$$\log \mu_{jk} \quad j = 1, \dots, 5, k = 1, 2, 3$$

Two models are fitted:

- Constant in dose level: $\log \mu_{jk} = \beta_{0k}$, for $k = 1, 2, 3$
- Varying with dose level $\log \mu_{jk} = \beta_{0k} + \beta_{1k}x_j$ for $j = 1, \dots, 5, k = 1, 2, 3$

Using the results from lectures, we can estimate the multinomial parameters by fitting this Poisson model, and then reparameterizing. This is achieved by the `multinom` function in the `nnet` library.

```
1 library(nnet)
2 Malform.x<-c(0,15,1,281,62.5,17,0,225,125,22,7,283,250,38,59,202,500,144,132,9)
3 Malform<-matrix(Malform.x,ncol=4,byrow=T)
4 Malform.X<-Malform[,1];Malform.Counts<-Malform[,2:4]
5 summary(multinom(Malform.Counts~1));summary(multinom(Malform.Counts~Malform.X))
```

The function `multinom` returns the coefficients and standard errors for the logit transformed parameters for each of the non-baseline categories. The results are as follows:

```
1 > summary(multinom(Malform.Counts~1))
2 # weights: 6 (2 variable)
3 initial value 1576.508634
4 final value 1180.313142
5 converged
6 Call: multinom(formula = Malform.Counts ~ 1)
7
8 Coefficients:
9 (Intercept)
10 2 -0.1705718
11 3 1.4437684
12
13 Std. Errors:
14 (Intercept)
15 2 0.09623711
16 3 0.07236605
17
18 Residual Deviance: 2360.626
19 AIC: 2364.626
20
```

```

21
22
23 > summary(multinom(Malform.Counts~Malform.X) )
24 # weights:  9 (4 variable)
25 initial value 1576.508634
26 iter  10 value 753.236516
27 final value 753.220332
28 converged
29 Call:  multinom(formula = Malform.Counts ~ Malform.X)
30
31 Coefficients:
32   (Intercept)      Malform.X
33 2    -0.983336   0.002100680
34 3     3.969451 -0.011908910
35
36 Std. Errors:
37   (Intercept)      Malform.X
38 2    0.2706013 0.0006455005
39 3    0.1909845 0.0006958429
40
41 Residual Deviance: 1506.441
42 AIC: 1514.441

```

In the first analysis, we have (lines 10,11)

$$\log \left(\frac{\hat{\pi}_2}{\hat{\pi}_1} \right) = \log \left(\frac{\hat{\pi}_2}{1 - \hat{\pi}_2 - \hat{\pi}_3} \right) = -0.1705718$$

and

$$\log \left(\frac{\hat{\pi}_3}{\hat{\pi}_1} \right) = \log \left(\frac{\hat{\pi}_3}{1 - \hat{\pi}_2 - \hat{\pi}_3} \right) = 1.4437684$$

whereas in the second analysis (lines 33,34)

$$\log \left(\frac{\hat{\pi}_2(x)}{\hat{\pi}_1(x)} \right) = \log \left(\frac{\hat{\pi}_2(x)}{1 - \hat{\pi}_2(x) - \hat{\pi}_3(x)} \right) = -0.983336 + 0.002100680x$$

and

$$\log \left(\frac{\hat{\pi}_3(x)}{\hat{\pi}_1(x)} \right) = \log \left(\frac{\hat{\pi}_3(x)}{1 - \hat{\pi}_2(x) - \hat{\pi}_3(x)} \right) = 1.4437684 = 3.969451 - 0.011908910x$$

The deviance comparison for this model is valid: we have nested models, and

- Constant in dose level: Deviance = 2360.626, $p = 2$
- Varying with dose level: Deviance = 1506.441, $p = 4$

and so Δ Deviance = 854.185 with Δ d.f. = 2.

4.4 OVERDISPERSION FOR BINOMIAL DATA

QUASI-BINOMIAL MODELS

The quasi-binomial model assumes a mean-variance structure of the form

$$\begin{aligned}\mathbb{E}[Y_i] &= m_i \pi_i \\ \text{Var}[Y_i] &= \phi m_i \pi_i (1 - \pi_i).\end{aligned}$$

for $i = 1, \dots, n$. Setting $Y_i^* = Y_i/m_i$, we have

$$\begin{aligned}\mathbb{E}[Y_i^*] &= \pi_i \\ \text{Var}[Y_i^*] &= \phi w_i \pi_i (1 - \pi_i) = \phi w_i v(\pi_i)\end{aligned}$$

say, where $w_i = 1/m_i$. Under the logistic link, where

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = x_i \beta = \eta_i$$

we have $g(t) = \log(t/(1 - t))$, $\dot{g}(t) = 1/(t(1 - t))$, so that the quasi-likelihood estimating equation is, as for the binomial case (see previous notes, and the handout on Iteratively Reweighted Least Squares)

$$\sum_{i=1}^n \frac{1}{w_i \dot{g}(\pi_i) v(\pi_i)} x_i^\top (y_i^* - \pi_i) = \sum_{i=1}^n m_i (y_i^* - \pi_i) x_i^\top = \sum_{i=1}^n (y_i - m_i \pi_i) x_i^\top = 0_p$$

If β contains p parameters, the estimate of the dispersion parameter is

$$\hat{\phi} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i^* - \hat{\pi}_i)^2}{w_i V(\hat{\pi}_i)} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i^* - \hat{\pi}_i)^2}{[\hat{\pi}_i(1 - \hat{\pi}_i)/m_i]} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i - m_i \hat{\pi}_i)^2}{m_i \hat{\pi}_i (1 - \hat{\pi}_i)}$$

In the standard Binomial model $\phi = 1$, so if $\hat{\phi} > 1$, this is evidence of overdispersion.

If $\hat{\phi} > 1$, then the standard errors of the estimates of the coefficients will be inflated, as the following example indicates. To generate overdispersion, set $n = 100$, $m_i = 20$, then

- $\pi_i \sim \text{Beta}(2, 2)$
- $Y_i | \pi_i \sim \text{Binomial}(m_i, \pi_i)$

and then fit the constant response models.

```
1 n<-100
2 pi<-rbeta(n, 2, 2)
3 m<-20
4 Y<-rbinom(n, m, pi)
5
6 summary(glm(cbind(Y, m-Y)~1, family=binomial))
7 summary(glm(cbind(Y, m-Y)~1, family=quasibinomial))
```

The results are as follows:

```
1 > summary(glm(cbind(Y,m-Y)~1,family=binomial))
2
3 Call: glm(formula = cbind(Y, m - Y) ~ 1, family = binomial)
4
5 Deviance Residuals:
6      Min       1Q   Median       3Q      Max
7 -5.2276  -1.3077   0.0447   1.3964   4.4886
8
9 Coefficients:
10             Estimate Std. Error z value Pr(>|z|)
11 (Intercept) -0.02000    0.04472  -0.447   0.655
12
13 (Dispersion parameter for binomial family taken to be 1)
14
15 Null deviance: 445.06  on 99  degrees of freedom
16 Residual deviance: 445.06  on 99  degrees of freedom
17 AIC: 765.77
18
19 Number of Fisher Scoring iterations: 3
20
21
22 > summary(glm(cbind(Y,m-Y)~1,family=quasibinomial))
23
24 Call: glm(formula = cbind(Y, m - Y) ~ 1, family = quasibinomial)
25
26 Deviance Residuals:
27      Min       1Q   Median       3Q      Max
28 -5.2276  -1.3077   0.0447   1.3964   4.4886
29
30 Coefficients:
31             Estimate Std. Error t value Pr(>|t|)
32 (Intercept) -0.02000    0.08988  -0.223   0.824
33
34 (Dispersion parameter for quasibinomial family taken to be 4.038788)
35
36 Null deviance: 445.06  on 99  degrees of freedom
37 Residual deviance: 445.06  on 99  degrees of freedom
38 AIC: NA
```

Note that the standard error doubles from 0.04472 (line 11) to 0.08988 (line 32), and that the estimate of the dispersion is $\hat{\phi} = 4.0388$; indeed, the standard error increases by $\sqrt{\hat{\phi}}$.

4.5 MULTINOMIAL RESPONSES

PROPORTIONAL ODDS MODEL

The proportional odds model is used for ordered multicategory response data; the logistic version of the model is specified for ungrouped data - that is, where $Y \sim Multinomial(1; \pi_1, \dots, \pi_K)$ as

$$\log \left(\frac{\Pr[Y \leq k|x]}{1 - \Pr[Y \leq k|x]} \right) = \beta_{0k} + x\beta_1 \quad k = 1, \dots, K-1$$

for predictor x and parameters β_1 . With the logistic link, for two values of the continuous covariate, x_A and x_B ,

$$\log \left(\frac{\Pr[Y \leq k|x_A]}{1 - \Pr[Y \leq k|x_A]} \right) - \log \left(\frac{\Pr[Y \leq k|x_B]}{1 - \Pr[Y \leq k|x_B]} \right) = \log \left(\frac{\Pr[Y \leq k|x_A]/\Pr[Y > |x_A]}{\Pr[Y \leq k|x_B]/\Pr[Y > |x_B]} \right) = (x_A - x_B)\beta_1$$

The quantity

$$\frac{\Pr[Y \leq k|x_A]/\Pr[Y > |x_A]}{\Pr[Y \leq k|x_B]/\Pr[Y > |x_B]}$$

is termed a *cumulative odds ratio*. For ordered categorical responses, the intercepts β_{0k} should be constrained to be increasing.

Example: Mental Impairment: A study of mental impairment in a sample of 40 adults examined the relationship between the mental impairment outcome (1-Well, 2-Mild, 3-Moderate, 4-Impaired) and two predictors x_1 (LE, number of life events during study period, continuous) and x_2 (SES, socioeconomic status, binary factor).

ID	Outcome Y	SES x_2	LE x_1	ID	Outcome Y	SES x_2	LE x_1
1	Well	1	1	21	Mild	1	9
2	Well	1	9	22	Mild	0	3
3	Well	1	4	23	Mild	1	3
4	Well	1	3	24	Mild	1	1
5	Well	0	2	25	Moderate	0	0
6	Well	1	0	26	Moderate	1	4
7	Well	0	1	27	Moderate	0	3
8	Well	1	3	28	Moderate	0	9
9	Well	1	3	29	Moderate	1	6
10	Well	1	7	30	Moderate	0	4
11	Well	0	1	31	Moderate	0	3
12	Well	0	2	32	Impaired	1	8
13	Mild	1	5	33	Impaired	1	2
14	Mild	0	6	34	Impaired	1	7
15	Mild	1	3	35	Impaired	0	5
16	Mild	0	1	36	Impaired	0	4
17	Mild	1	8	37	Impaired	0	4
18	Mild	1	2	38	Impaired	1	8
19	Mild	0	5	39	Impaired	0	8
20	Mild	1	5	40	Impaired	0	9

The outcome levels are of increasing severity of impairment. These data are available from the course website. The model proposed is

$$\log \left(\frac{\Pr[Y \leq k|x]}{1 - \Pr[Y \leq k|x]} \right) = \beta_{0k} + \beta_1 x_1 + \beta_2 \mathbb{1}_{\{1\}} x_2 \quad k = 1, 2, 3$$

where the constraint $\beta_{01} < \beta_{02} < \beta_{03}$ respects the ordered nature of the outcomes. The parameter β_2 measures the effect (on the linear predictor scale) of socio-economic status.

The proportional odds model can be fitted with `polr` function from the MASS library:

```
1 >Mental<-read.table("MentalImpairment.txt",header=F)
2 >Mental1<-Mental[,1:4];>Mental2<-Mental[,5:8];names(Mental2)<-names(Mental1)
3 >Mental<-rbind.data.frame(Mental1,Mental2)
4 >names(Mental)<-c("ID","Y","ses","life")
5 >Mental$ses<-as.factor(Mental$ses)
6 >Mental$Y<-as.factor(rep(c(1,2,3,4),c(12,12,7,9)))
7 >summary(polr(Y~life+ses,data=Mental))
8 Call: polr(formula = Y ~ life + ses, data = Mental)
9 Coefficients:
10      Value Std. Error t value
11 life  0.3189    0.1210   2.635
12 ses1 -1.1112    0.6109  -1.819
13
14 Intercepts:
15      Value Std. Error t value
16 1|2 -0.2819    0.6423  -0.4389
17 2|3  1.2128    0.6607   1.8357
18 3|4  2.2094    0.7210   3.0644
19
20 Residual Deviance: 99.0979
21 AIC: 109.0979
```

Note: the `polr` function reports parameter estimates for the coefficients for non-intercept terms on the negative scale, that is, it reports $-\hat{\beta}_1$ and $-\hat{\beta}_2$. Estimated coefficients with standard errors from lines 5-6 are $\hat{\beta}_1 = -0.3189$ (0.1210), $\hat{\beta}_2 = 1.1112$ (0.6109). The intercepts $\beta_{01} < \beta_{02} < \beta_{03}$ are estimated as -0.2819, 1.2128, 2.2094 (lines 10-12). Using a chi-squared test we can assess the importance of the variables in the model.

```
1 > drop1(fit.mental.polr,test="Chisq")
2 Single term deletions
3
4 Model: Y ~ life + ses
5      Df      AIC      LRT Pr(>Chi)
6 <none>    109.10
7 life    1 114.87  7.7765 0.005293 **
8 ses     1 110.53  3.4292 0.064054 .
9 ---
10 Signif. codes:  0   ***  0.001  **  0.01  *  0.05  .  0.1    1
```

It is evident that the `ses` predictor is not influential, so this variable can be removed:

```
1 > summary(polr(Y~life,data=Mental))
2 Call: polr(formula = Y ~ life, data = Mental)
3 Coefficients:
4      Value Std. Error t value
5 life 0.2879    0.1175   2.451
6
7 Intercepts:
8      Value Std. Error t value
9 1|2 0.2614 0.5639    0.4636
```

```

10 2|3 1.6563 0.6171      2.6838
11 3|4 2.5876 0.6938      3.7296
12
13 Residual Deviance: 102.5271
14 AIC: 110.5271

```

The model can also be fitted using the `vglm` function from the VGAM library.

```

1 >library(VGAM)
2 >fit.mental.vglm<-vglm(Y~life+ses,data=Mental,propodds(reverse=F))
3 >summary(fit.mental.vglm)
4 Call: vglm(formula = Y~life + ses, family = propodds, data = Mental)
5 Pearson Residuals:
6               Min           1Q       Median           3Q          Max
7 logit(P[Y< = 1]) -1.5687 -0.70508 -0.21026  0.80709  2.7155
8 logit(P[Y< = 2]) -2.3297 -0.46662  0.26578  0.69049  1.6148
9 logit(P[Y< = 3]) -3.6874  0.11976  0.20398  0.41914  1.8931
10
11 Coefficients:
12             Estimate Std. Error  z value
13 (Intercept):1 -0.28176      0.62304 -0.45223
14 (Intercept):2  1.21291      0.65119  1.86260
15 (Intercept):3  2.20947      0.71719  3.08075
16 life          -0.31888      0.11944 -2.66973
17 ses1           1.11112      0.61427  1.80884
18
19 Number of linear predictors: 3
20 Names of linear predictors: logit(P[Y< = 1]), logit(P[Y< = 2]), logit(P[Y< = 3])
21
22 Dispersion Parameter for cumulative family: 1
23 Residual deviance: 99.0979 on 115 degrees of freedom
24 Log-likelihood: -49.54895 on 115 degrees of freedom

```

Using the `propodds` argument without the `reverse=T` option computes via the probabilities

$$\log\left(\frac{\Pr[Y \geq k|x]}{1 - \Pr[Y \geq k|x]}\right) = \beta_{0k} + \beta_1 x_1 + \beta_2 1_{\{1\}} x_2 \quad k = 1, 2, 3$$

```

1 > fit.mental.vglm<-vglm(Y~life+ses,data=Mental,propodds)
2 > summary(fit.mental.vglm)
3 Call: vglm(formula = Y ~ life + ses, family = propodds, data = Mental)
4 Pearson Residuals:
5               Min           1Q       Median           3Q          Max
6 logit(P[Y> = 2]) -2.7155 -0.80709  0.21026  0.70508  1.5687
7 logit(P[Y> = 3]) -1.6148 -0.69049 -0.26578  0.46662  2.3297
8 logit(P[Y> = 4]) -1.8931 -0.41914 -0.20398 -0.11976  3.6874
9
10 Coefficients:
11             Estimate Std. Error  z value
12 (Intercept):1  0.28176      0.62304  0.45223
13 (Intercept):2 -1.21291      0.65119 -1.86260
14 (Intercept):3 -2.20947      0.71719 -3.08075
15 life           0.31888      0.11944  2.66973
16 ses1          -1.11112      0.61427 -1.80884

```

```

17
18 Number of linear predictors: 3
19 Names of linear predictors: logit(P[Y> = 2]), logit(P[Y> = 3]), logit(P[Y> = 4])
20
21 Dispersion Parameter for cumulative family: 1
22 Residual deviance: 99.0979 on 115 degrees of freedom
23 Log-likelihood: -49.54895 on 115 degrees of freedom

```

The multinom function could also be used for these data to fit a conventional multicategory model:

```

1 >fit.mental.multinom<-multinom(Y ~life + ses, data = Mental)
2 >summary(it.mental.multinom)
3
4 Call: multinom(formula = Y ~ life + ses, data = Mental)
5 Coefficients:
6   (Intercept)      life      ses1
7 2  -0.6552117  0.2285365 -0.2487293
8 3  -0.6621855  0.2930040 -1.9437361
9 4  -1.9025935  0.5601587 -1.6062505
10
11 Std. Errors:
12   (Intercept)      life      ses1
13 2    0.8853634  0.1832107  0.9047728
14 3    0.9582681  0.2199351  1.1081671
15 4    1.1107458  0.2190388  1.0804573
16
17 Residual Deviance: 96.69826
18 AIC: 114.6983

```

The estimated probabilities arising from the two models are broadly similar, but different:

ID	Prop. Odds				Multinomial			
	1	2	3	4	1	2	3	4
1	0.625	0.256	0.071	0.047	0.602	0.307	0.060	0.032
2	0.115	0.252	0.244	0.389	0.102	0.322	0.105	0.471
3	0.390	0.350	0.145	0.115	0.395	0.399	0.094	0.111
4	0.468	0.329	0.117	0.086	0.467	0.375	0.083	0.075
5	0.285	0.355	0.188	0.172	0.312	0.256	0.289	0.143
6	0.696	0.215	0.054	0.035	0.663	0.268	0.049	0.020
7	0.354	0.356	0.159	0.131	0.384	0.251	0.265	0.100
8	0.468	0.329	0.117	0.086	0.467	0.375	0.083	0.075
9	0.468	0.329	0.117	0.086	0.467	0.375	0.083	0.075
10	0.197	0.326	0.225	0.252	0.196	0.394	0.113	0.297
11	0.354	0.356	0.159	0.131	0.384	0.251	0.265	0.100
12	0.285	0.355	0.188	0.172	0.312	0.256	0.289	0.143
13	0.318	0.357	0.174	0.151	0.324	0.412	0.104	0.160
14	0.100	0.232	0.242	0.426	0.097	0.198	0.289	0.416
15	0.468	0.329	0.117	0.086	0.467	0.375	0.083	0.075
:	:	:	:	:	:	:	:	:

4.6 CONDITIONAL LOGISTIC REGRESSION

The standard logistic regression model for binary responses sets

$$\pi_i(x_i) = \Pr[Y_i = y_i | x_i] = \frac{\exp\{y_i(\beta_0 + x_i\beta)\}}{1 + \exp\{(\beta_0 + x_i\beta)\}} \quad y_i = 0, 1.$$

Computing the likelihood, $T_1(\mathbf{Y})$ (scalar) is sufficient for β_0 , and $T_2(\mathbf{Y}, \mathbf{X})$ ($1 \times p$) is sufficient for β

$$L_n(\beta_0, \beta) = \frac{\exp\{T_1(\mathbf{y})\beta_0 + T_2(\mathbf{y}, \mathbf{x})\beta\}}{\prod_{i=1}^n \{1 + \exp\{(\beta_0 + x_i\beta)\}\}} \quad \text{where} \quad T_1(\mathbf{y}) = \sum_{i=1}^n y_i \quad T_2(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^n y_i x_i$$

Here, β_0 is a *nuisance* parameter; we are interested in the impact of the covariates on response, so β is the *parameter of interest*. Thus, consider the likelihood conditional on $T_1(\mathbf{Y}) = t_1$:

$$\Pr[\mathbf{Y} = \mathbf{y} | T_1 = t_1] = \frac{\Pr[Y_1 = y_1, \dots, Y_n = y_n]}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \Pr[Y_1 = z_1, \dots, Y_n = z_n]}$$

where $\mathcal{Y}(t) = \{(z_1, \dots, z_n) : T_1(\mathbf{z}) = t\}$. We have from above that

$$\Pr[Y_1 = y_1, \dots, Y_n = y_n] \equiv L_n(\beta_0, \beta) = \frac{\exp\{T_1(\mathbf{y})\beta_0 + T_2(\mathbf{y}, \mathbf{x})\beta\}}{\prod_{i=1}^n \{1 + \exp\{(\beta_0 + x_i\beta)\}\}} = \frac{\exp\{t_1\beta_0 + T_2(\mathbf{y}, \mathbf{x})\beta\}}{\prod_{i=1}^n \{1 + \exp\{(\beta_0 + x_i\beta)\}\}}$$

where, note, the denominator does not depend on $\mathbf{y}$. Therefore

$$\Pr[\mathbf{Y} = \mathbf{y} | T_1 = t_1] = \frac{\exp\{t_1\beta_0 + T_2(\mathbf{y}, \mathbf{x})\beta\} / \left\{ \prod_{i=1}^n \{1 + \exp\{(\beta_0 + x_i\beta)\}\} \right\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{z_1\beta_0 + T_2(\mathbf{z}, \mathbf{x})\beta\} / \left\{ \prod_{i=1}^n \{1 + \exp\{(\beta_0 + x_i\beta)\}\} \right\}} = \frac{\exp\{T_2(\mathbf{y}, \mathbf{x})\beta\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{T_2(\mathbf{z}, \mathbf{x})\beta\}}$$

This conditional distribution does not involve β_0 . Therefore, we can consider the *conditional likelihood*

$$L_n^{(T_1)}(\beta; t_1) = \frac{\exp\{T_2(\mathbf{y}, \mathbf{x})\beta\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1)} \exp\{T_2(\mathbf{z}, \mathbf{x})\beta\}} \quad (11)$$

and estimate β using it. This is a regular likelihood, so ML estimators have standard properties.

This argument can be taken further; suppose $\beta = (\beta_1, \dots, \beta_p)^\top$, but that only β_1 is of interest. We can form the conditional likelihood conditioning on $T_1 = t_1$ **and** $T_{2j} = t_{2j}$ for $j = 2, \dots, p$, where T_{2j} is the j th element of $T_2(\mathbf{y}, \mathbf{x})$. This leads to the following conditional likelihood for β_1 :

$$L_n^{(T_1, T_{22}, \dots, T_{2p})}(\beta_1; t_1, t_{22}, \dots, t_{2p}) = \frac{\exp\{T_{21}(\mathbf{y}, \mathbf{x})\beta_1\}}{\sum_{\mathbf{z} \in \mathcal{Y}(t_1, t_{22}, \dots, t_{2p})} \exp\{T_{21}(\mathbf{y}, \mathbf{z})\beta_1\}} \quad (12)$$

where $\mathcal{Y}(t_1, t_{22}, \dots, t_{2p}) = \{(z_1, \dots, z_n) : T_1(\mathbf{z}) = t_1, T_{2j}(\mathbf{z}, \mathbf{x}) = t_{2j}, j = 2, \dots, p\}$. Again, conditional ML estimation of β_1 is relatively straightforward. Notice that testing the hypotheses $H_0 : \beta = 0_p$ using (11), or $H_0 : \beta_1 = 0$ using (12), all that is required is the probability distribution of the test statistics

$$L_n^{(T_1)}(0_p; t_1) \quad \text{and} \quad L_n^{(T_1, T_{22}, \dots, T_{2p})}(0; t_1, t_{22}, \dots, t_{2p}).$$

These are discrete distributions, so to compute p -values, we must compute the proportion of possible configurations of $\mathbf{y} = (y_1, \dots, y_n)$ yielding test statistics at least as large as the observed values.

4.7 MATCHED PAIRS ANALYSIS

A *Matched Pair* study involves the measurement of pairs of responses; the pairing can be the result of making two observations on the same experimental unit (at time 1 and time 2, or before and after treatment etc.), or on matched (by age, sex or other characteristics) pairs of different experimental units.

Consider $2n$ data that collected as matched pairs, that is n pairs of data (y_{i1}, y_{i2}) for $i = 1, \dots, n$. Suppose that the outcomes are binary, so that the data can be collated into a 2×2 table:

		$j = 2$	
		$Y = 1$	$Y = 0$
$j = 1$	$Y = 1$	n_{11}	n_{10}
	$Y = 0$	n_{01}	n_{00}

where n_{11} counts the number of pairs in which $y_{i1} = 1, y_{i2} = 1$ etc. Consider a multinomial model for these data, $(N_{00}, N_{01}, N_{10}, N_{11}) \sim \text{Multinomial}(n; \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. The ML estimates for these parameters are

$$(\hat{\pi}_{00}, \hat{\pi}_{01}, \hat{\pi}_{10}, \hat{\pi}_{11}) = \frac{1}{n}(n_{00}, n_{01}, n_{10}, n_{11})$$

Denote by π_{1*} and π_{*1} the probabilities of response $Y = 1$ for $j = 1$ and $j = 2$ respectively, that is

$$\pi_{1*} = \pi_{10} + \pi_{11} \quad \pi_{*1} = \pi_{01} + \pi_{11}$$

- (a) We may estimate (π_{1*}, π_{*1}) from the observed data, but the estimators will be **dependent** due to pairing; in fact, by properties of the multinomial distribution,

$$\text{Cov}[\hat{\pi}_{1*}, \hat{\pi}_{*1}] = \text{Cov}[\hat{\pi}_{10} + \hat{\pi}_{11}, \hat{\pi}_{01} + \hat{\pi}_{11}] = \frac{(\pi_{00}\pi_{11} - \pi_{01}\pi_{10})}{n}$$

- (b) If $\pi_{1*} = \pi_{*1}$, so that $\pi_{0*} = \pi_{*0}$ also, then the marginal distributions are identical (so we have *marginal homogeneity*). Furthermore,

$$\delta = \pi_{1*} - \pi_{*1} = \pi_{10} - \pi_{01}$$

so if $\delta = 0$, $\pi_{1*} = \pi_{*1}$, then $\pi_{10} = \pi_{01}$, and this corresponds to *symmetry* in the 2×2 table.

- (c) We have that $\hat{\delta} = \hat{\pi}_{1*} - \hat{\pi}_{*1}$, and, by the multivariate Delta method,

$$\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} \mathcal{N}(0, V(\pi))$$

where

$$V(\pi) = \pi_{1*}(1 - \pi_{1*}) + \pi_{*1}(1 - \pi_{*1}) - 2(\pi_{00}\pi_{11} - \pi_{01}\pi_{10})$$

- (d) For large n , $\hat{\delta}$ is approximately distributed as $\mathcal{N}(\delta, V(\hat{\pi})/n)$, where

$$V(\hat{\pi}) = (\hat{\pi}_{10} + \hat{\pi}_{01}) - (\hat{\pi}_{10} - \hat{\pi}_{01})^2$$

- (e) Under the null hypothesis of $H_0 : \delta = 0$, or equivalently $H_0 : \pi_{1*} = \pi_{*1}$ or $H_0 : \pi_{10} = \pi_{01}$ (corresponding to marginal homogeneity), the Wald test statistic W_n takes the form

$$W_n = \frac{\sqrt{n} \hat{\delta}}{\sqrt{V(\hat{\pi})}}$$

- (f) Alternatively, a different (simpler) estimator of the asymptotic variance is $V_0(\hat{\pi}) = (\hat{\pi}_{10} + \hat{\pi}_{01})$. Then

$$W_n = \frac{\sqrt{n} \hat{\delta}}{\sqrt{V_0(\hat{\pi})}} = \frac{n_{10} - n_{01}}{\sqrt{n_{01} + n_{10}}}$$

The squared Wald statistic, W_n^2 may be used; under H_0 , $W_n^2 \sim \chi_1^2$.

Example:- Prime Minister Approval 1600 UK voters were asked on two occasions separated by six months whether they approved or disapproved of the performance of the Prime Minister.

		Second Survey	
		Approve	Disapprove
First Survey	Approve	794	150
	Disapprove	86	570

Here $\hat{\delta} = \hat{\pi}_{1*} - \hat{\pi}_{*1} = (150 + 794)/1600 - (86 + 794)/1600 = 0.04$, with estimated standard error

$$\frac{1}{\sqrt{n}} \sqrt{(\hat{\pi}_{10} + \hat{\pi}_{01}) - (\hat{\pi}_{10} - \hat{\pi}_{01})^2} = \frac{1}{\sqrt{1600}} \sqrt{\left(\frac{150}{1600} + \frac{86}{1600}\right) - \left(\frac{150}{1600} - \frac{86}{1600}\right)^2} = 0.009549215$$

Thus the Wald statistic is

$$W_n = \frac{0.04}{0.00955} = 4.188$$

which is a highly significant result (after comparison with the standard Normal distribution); the probability of approval has decreased significantly. The second Wald statistic uses $V_0(\hat{\pi})$

$$W_n = \frac{n_{10} - n_{01}}{\sqrt{n_{01} + n_{10}}} = \frac{150 - 86}{\sqrt{150 + 86}} = 4.166$$

Matched pairs can also be analyzed using *conditional logistic regression*; a pair or *stratum* indicator is used to identify paired observations, and then the comparisons made between surveys which are treated as a levels of a binary factor. In R, conditional logistic regression can be carried out using the `clogit` function in the `survival` library. In this analysis X is the survey number (1-first survey, 0-second survey), and Y is approval indicator (1-approve, 0-disapprove):

```
1 library(survival)
2 Pair<-rep(c(1:1600),2)
3 X<-rep(c(1,0),each=1600)
4 Y<-rep(0,3200)
5 for(i in 1:794){Y[Pair == i]<-1}
6 for(i in 795:944){Y[Pair == i]<-c(1,0)}
7 for(i in 945:1030){Y[Pair == i]<-c(0,1)}
8 for(i in 1030:1600){Y[Pair == i]<-0}
9 Prime<-data.frame(X,Y,Pair)
10 clogit(Y~X+strata(Pair),data=Prime)
11
12
13 > clogit(Y~X+strata(Pair),data=Prime)
14 Call: clogit(Y ~ X + strata(Pair), data = Prime)
15
16      coef exp(coef) se(coef)      z      p
17 X 0.568      1.76      0.136 4.18 2.9e-05
18
19 Likelihood ratio test=18.2 on 1 df, p=1.97e-05 n= 3200, number of events= 1823
```

The `clogit` function on line 10 fits the conditional logistic regression. The results on line 17 indicate a significant association; the estimate 0.568 is the subject-specific log odds ratio, $\hat{\zeta}$, which is significantly less than zero (p -value is 2.90e-5). This confirms the result that approval rating decreased between surveys.

Conditional Logistic Regression for Matched Pairs or Stratified Data.

Let (Y_{i1}, Y_{i2}) be the responses for the i th matched pair, and suppose that $\pi_{ij} = \Pr[Y_{ij} = 1]$. Let $x_j = j - 1$, for $j = 1, 2$, and consider the model

$$\log \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_{0i} + \zeta x_j = \begin{cases} \beta_{0i} & j = 1 \\ \beta_{0i} + \zeta & j = 2 \end{cases}$$

The parameters of the resulting model are the *pair-* or *subject-specific* intercepts, $\beta_{01}, \dots, \beta_{0n}$, and the *subject-specific* log odds ratio, ζ .

Note that the interpretation of ζ is different from the interpretation of the *population* log odds ratio β_1 ($= \log \psi = \xi$ in previous notation) in the analysis of 2×2 tables for independent data using logistic regression; β measures association at the population level, and arises as a population average across independent individuals, whereas ζ measures association at the individual matched pair level.

We can compute the full likelihood for these data as

$$L_n(\beta_{01}, \dots, \beta_{0n}, \zeta) = \prod_{i=1}^n \left\{ \left(\frac{e^{\beta_{0i}}}{1 + e^{\beta_{0i}}} \right)^{y_{i1}} \left(\frac{1}{1 + e^{\beta_{0i}}} \right)^{1-y_{i1}} \left(\frac{e^{\beta_{0i}+\zeta}}{1 + e^{\beta_{0i}+\zeta}} \right)^{y_{i2}} \left(\frac{1}{1 + e^{\beta_{0i}+\zeta}} \right)^{1-y_{i2}} \right\}.$$

Then following the argument above, we can form the conditional likelihood to eliminate the nuisance parameters $\beta_{01}, \dots, \beta_{0n}$, that is, we condition on $s_i = y_{i1} + y_{i2}$ for $i = 1, \dots, n$. The random variables $S_1, \dots, S_n$ each can take the values 0, 1, 2. It can be easily shown that

$$\Pr[Y_{i1} = y_{i1}, Y_{i2} = y_{i2} | S_i = s] = \begin{cases} 1 & (s, y_{i1}, y_{i2}) = (0, 0, 0) \text{ or } (2, 1, 1) \\ \frac{e^{y_{i2}\zeta}}{1 + e^\zeta} & (s, y_{i1}, y_{i2}) = (1, 0, 1) \text{ or } (1, 1, 0) \end{cases}$$

and therefore the conditional likelihood for ζ , given $\mathbf{S} = (s_1, \dots, s_n)$ is

$$L_n^{(\mathbf{S})}(\zeta; s_1, \dots, s_n) = \frac{e^{n_{01}\zeta}}{(1 + e^\zeta)^{n_{01}+n_{10}}}$$

where n_{01} is the number of observations with $y_{i1} = 0, y_{i2} = 1$, and n_{10} is the number of observations with $y_{i1} = 1, y_{i2} = 0$ (that is, the *discordant* pairs). It is easy to derive the ML estimate, and the Delta method gives the estimated standard error:

$$\hat{\zeta} = \log \left(\frac{n_{01}}{n_{10}} \right) \quad \text{e.s.e}(\hat{\zeta}) = \sqrt{\frac{1}{n_{01}} + \frac{1}{n_{10}}}$$

whereas the population estimate is $\hat{\xi} = \log(n_{00}n_{11}/n_{01}n_{10})$

This approach can be extended to handle *stratified* data, where each of K strata has its own intercept parameter but stratum k contains n_k individuals rather than a single matched pair,

$$\log \left(\frac{\pi_{ijk}}{1 - \pi_{ijk}} \right) = \beta_{0k} + \zeta x_j \quad i = 1, \dots, n_k, j = 1, 2, k = 1, \dots, K$$

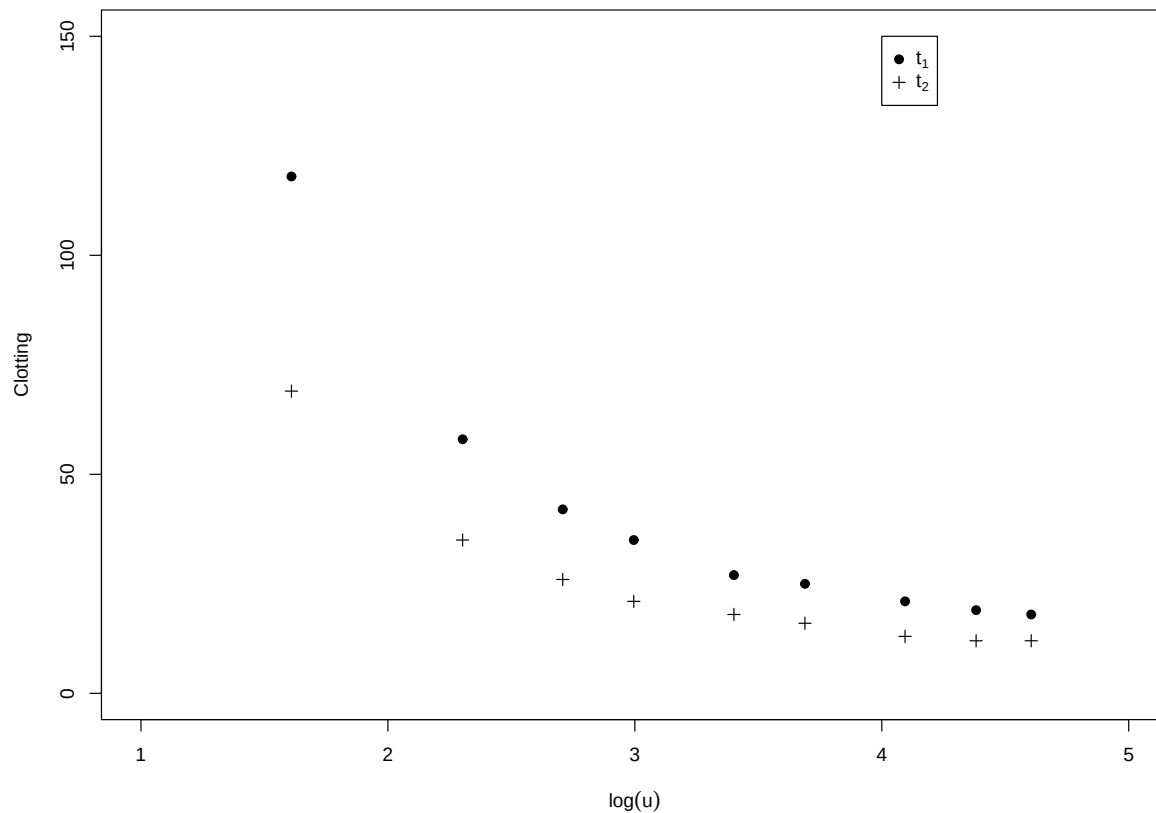
and problems with multiple covariates.

5 SPECIAL TOPICS

5.1 GAMMA GLM

Example:- Blood clotting time The following data representing the clotting times of two blood samples at nine different dilutions u . The first sample is from a healthy subject, the second is from a hospital patient.

u	5	10	15	20	30	40	60	80	100
t_1	118	58	42	35	27	25	21	19	18
t_2	69	35	26	21	18	16	13	12	12



Six models were fitted with different link functions and linear predictors: let $x = \log(u)$.

Mod 1 Inverse link: $\eta = \beta_0 + \beta_1 x$.

Mod 2 Log link: $\eta = \beta_0 + \beta_1 x$.

Mod 3 Inverse link: $\eta = \beta_0 + \beta_1/x$.

Mod 4 Log link: $\eta = \beta_0 + \beta_1/x$.

Mod 5 Inverse link: $\eta = \beta_0 + \beta_1 x + \beta_2/x$.

Mod 6 Log link: $\eta = \beta_0 + \beta_1 x + \beta_2/x$.

```
1 clotting <- data.frame(u = c(5,10,15,20,30,40,60,80,100),
2   lot1 = c(118,58,42,35,27,25,21,19,18), lot2 = c(69,35,26,21,18,16,13,12,12))
3 mod1<-glm(lot1 ~ log(u), data=clotting, family=Gamma)
4 mod2<-glm(lot1 ~ log(u), data=clotting, family=Gamma(link=log))
```

```

5 mod3<-glm(lot1 ~ I(1/log(u)), data=clotting, family=Gamma)
6 mod4<-glm(lot1 ~ I(1/log(u)), data=clotting, family=Gamma(link=log))
7 mod5<-glm(lot1 ~ log(u)+I(1/log(u)), data=clotting, family=Gamma)
8 mod6<-glm(lot1 ~ log(u)+I(1/log(u)), data=clotting, family=Gamma(link=log))
9
10 > summary(mod5)
11
12 Call: glm(formula = lot1 ~ log(u) + I(1/log(u)), family = Gamma, data = clotting)
13
14 Deviance Residuals:
15      Min       1Q   Median       3Q      Max
16 -0.044739 -0.001576  0.000990  0.009777  0.020533
17
18 Coefficients:
19             Estimate Std. Error t value Pr(>|t|)
20 (Intercept) -0.036158   0.003574 -10.117 5.42e-05 ***
21 log(u)       0.019114   0.000710  26.922 1.73e-07 ***
22 I(1/log(u))  0.022286   0.004034   5.525 0.00148 **
23 ---
24 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
25
26 (Dispersion parameter for Gamma family taken to be 0.0004605922)
27
28 Null deviance: 3.512826 on 8 degrees of freedom
29 Residual deviance: 0.002813 on 6 degrees of freedom
30 AIC: 23.941
31
32 Number of Fisher Scoring iterations: 3
33
34
35 > summary(mod6)
36
37 Call: glm(formula = lot1 ~ log(u) + I(1/log(u)), family = Gamma(link = log),
38 data = clotting)
39
40 Deviance Residuals:
41      Min       1Q   Median       3Q      Max
42 -0.050709 -0.007464 -0.000478  0.009112  0.037492
43
44 Coefficients:
45             Estimate Std. Error t value Pr(>|t|)
46 (Intercept)  2.96483    0.18498  16.028 3.75e-06 ***
47 log(u)      -0.18148    0.03168  -5.729 0.00123 **
48 I(1/log(u))  3.39912    0.24573  13.833 8.89e-06 ***
49 ---
50 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1    1
51
52 (Dispersion parameter for Gamma family taken to be 0.0008025472)
53
54 Null deviance: 3.5128263 on 8 degrees of freedom
55 Residual deviance: 0.0048588 on 6 degrees of freedom
56 AIC: 28.860
57
58 Number of Fisher Scoring iterations: 3

```

RESULTS

```

1 X<-rep(log(clotting$u),2)
2 Y<-c(clotting$lot1,clotting$lot2)
3 L<-as.factor(rep(c(1,2),each=9))
4
5 mod.fit1<-glm(Y ~ L*X, family=Gamma(link=log))
6 mod.fit2<-glm(Y ~ L*I(1/X), family=Gamma(link=log))
7 mod.fit3<-glm(Y ~ L*(X+I(1/X)), family=Gamma(link=log))
8 mod.fit4<-glm(Y ~ L+(X+I(1/X)), family=Gamma(link=log))
9
10 > summary(mod.fit4)
11
12 Call: glm(formula = Y ~ L + (X + I(1/X)), family = Gamma(link = log))
13
14 Deviance Residuals:
15      Min       1Q   Median       3Q      Max
16 -0.054279 -0.019740  0.002937  0.017107  0.068725
17
18 Coefficients:
19             Estimate Std. Error t value Pr(>|t|)
20 (Intercept)   2.95755    0.16321  18.121 4.08e-11 ***
21 L2            -0.46966    0.01664 -28.221 9.69e-14 ***
22 X             -0.17257    0.02791  -6.182 2.38e-05 ***
23 I(1/X)         3.33325    0.21653  15.394 3.61e-10 ***
24 ---
25 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
26
27 (Dispersion parameter for Gamma family taken to be 0.001246303)
28
29 Null deviance: 7.708667  on 17  degrees of freedom
30 Residual deviance: 0.017350  on 14  degrees of freedom
31 AIC: 53.699
32
33 Number of Fisher Scoring iterations: 3

```

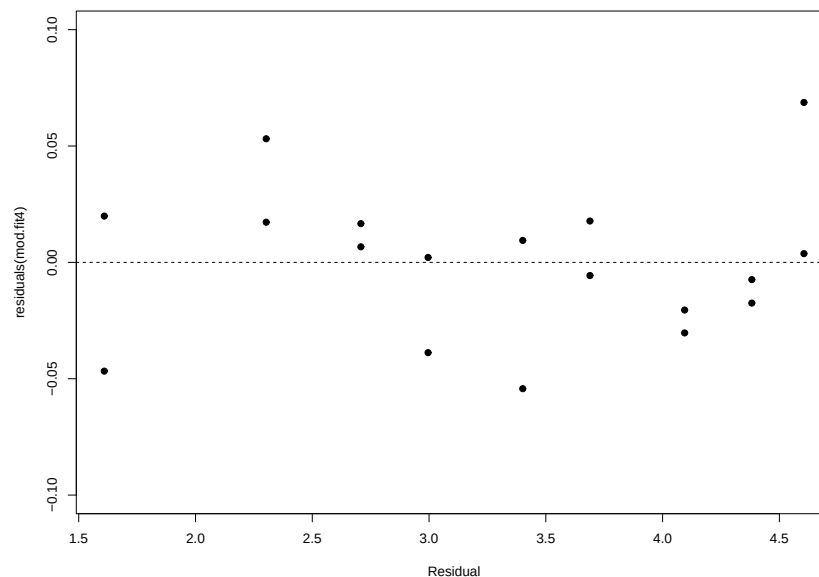


Table of the Chi-squared distribution

Entries in table are $\chi^2_{\alpha}(\nu)$: the α tail quantile of Chi-squared(ν) distribution
 α given in columns, ν given in rows.

ν	Left-tail					Right-tail				
	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

