An Introduction to Generalized Linear Models (third edition, 2008) by Annette Dobson & Adrian Barnett
Outline of solutions for selected exercises

CHAPTER 1

1.1 $\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim \mathbf{N}\left[\begin{pmatrix} 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 23 & 2 \\ 2 & 53 \end{pmatrix}\right]$.

1.2 (a) $\chi^2(1)$, by property 1 for the chi-squared distribution (page 7).

(b) $\mathbf{y}^{\mathrm{T}}\mathbf{y} = Y_1^2 + \left(\frac{Y_2 - 3}{2}\right)^2 \sim \chi^2(2)$, by property 3.

(c) $\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} = Y_1^2 + Y_2^2/4 \sim \chi^2(2, 9/4)$, by property 6.

1.3 (a) $Q_1 = (\mathbf{y}-\boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = \frac{1}{35}\left[9(Y_1-2)^2 - 2(Y_1-2)(Y_2-3) + 4(Y_2-3)^2\right] \sim \chi^2(2)$. This can be verified by considering $U = \frac{1}{2\sqrt{15}}(-3Y_1 + 2Y_2)$ and $V = \frac{1}{2\sqrt{21}}(3Y_1 + 2Y_2 - 12)$ and showing that U and V are independent, identically distributed random variables with the distribution $N(0, 1)$ and that $Q = U^2 + V^2$.

(b) $Q_2 = \mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} = \frac{1}{35}\left(9Y_1^2 - 2Y_1Y_2 + 4Y_2^2\right) \sim \chi^2(2, 12/7)$ By property 6.

This can be verified by showing that $Q_2 = U^2 + \left(V + \frac{2\sqrt{3}}{\sqrt{7}}\right)^2$, so by property 4 $\lambda = \left(\frac{2\sqrt{3}}{\sqrt{7}}\right)^2 = \frac{12}{7}$.

1.4 (a) $\bar{Y} \sim N\left(\mu, \sigma^2/n\right)$

(c) and (d) follow from results on p.10, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.

(e) If $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and $U^2 = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$

then $\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{Z}{\left[U^2/(n-1)\right]^{1/2}} \sim t(n-1)$.

1.5 (b) and (c) $\hat{\beta} = \log(\bar{y})$

1.6 (a) The proportions of females range from 0.368 to 0.621.

(b) $\hat{\theta} = \sum y_i / \sum n_i = 0.4946$

(c) Plot $l(\theta) = \log\theta \sum y_i + \log(1-\theta)\sum(n_i - y_i)$ (ignoring the constant term), against θ for various θ and find where the minimum value is.

CHAPTER 2

2.1 (a) There is little evidence of any difference between the two groups.

(b) Assuming equal variances in the two groups, the unpaired t-test of $\mu_1 = \mu_2$ against $\mu_1 \neq \mu_2$ gives t = 0.51, d.f. = 38, p-value = 0.613. This provides little evidence against the null hypothesis that the group means are equal.

(c) – (f) For these data $\hat{S}_0 = 26.2302, \hat{S}_1 = 26.0519$.

(g) F = 0.26 which is small compared to the distribution F (1, 38) so the data provide little evidence against H_0.

(h) $t^2 = 0.51^2 = 0.26 = F$. The conclusions are the same.

(i) The residuals are consistent with the assumptions of independence, equal variances and Normality.

2.2 (a) For an unpaired t-test, assuming equal variances, t = 0.64, d.f. = 38, p-value = 0.524 so there is little evidence against the null hypothesis of no effect. The 95% confidence interval for the difference in means, $\mu_2 - \mu_1$, is $(-5.68, 10.97)$.

(b) Let $E(D_k) = \mu_D = \mu_1 - \mu_2$. Then H_0 corresponds to the model. $E(D_k) = \mu_D = 0$; $D_k \sim N(\mu_D, \sigma_D^2)$ and H_1 corresponds to the model $E(D_k) = \mu_D$; $D_k \sim N(\mu_D, \sigma_D^2)$. The test of H_0 against H_1 yields $F = 8.24$ which is statistically significant when compared with the $F(1,19)$ distribution. The 95% confidence interval for μ_D is (0.718, 4.572) showing a statistically significant reduction in weight.

(c) The conclusions are different.

(d) For (a) it is assumed that the Y_{jk}'s are independent and $Y_{jk} \sim N(\mu_j, \sigma^2)$ for all j and for all k. For (b) it is assumed that the D_k's are independent with $D_k \sim N(\mu_D, \sigma_D^2)$. The analysis in (b) does not involve assuming that Y_{1k} and Y_{2k} (i.e., 'before' and 'after' weights for the same man) are independent, so it is more appropriate.

2.4 $\exp[E(\mathbf{y})] = \mathbf{X\beta}$ where $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ with

$$\mathbf{y} = \begin{bmatrix} 3.15 \\ 4.85 \\ 6.50 \\ 7.20 \\ 8.25 \\ 16.50 \end{bmatrix} \text{ and } \mathbf{X} = \begin{bmatrix} 1 & 1.0 & 1.00 \\ 1 & 1.2 & 1.44 \\ 1 & 1.4 & 1.96 \\ 1 & 1.6 & 2.56 \\ 1 & 1.8 & 3.24 \\ 1 & 2.0 & 4.00 \end{bmatrix}.$$

2.5 $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

CHAPTER 3

3.1 (a) Response: Y_i = weight – continuous scale, possibly Normally distributed;
Explanatory variables: x_{i1} = age, x_{i2} = sex (indicator variable), x_{i3} = height, x_{i4} = mean daily food intake, and x_{i5} = mean daily energy expenditure;
$\mathrm{E}(Y_i) = \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$; $Y_i \sim N(\mu_i, \sigma^2)$.

(b) Response: Y = number of mice infected in each group of $n = 20$ mice;
Explanatory variables: $x_{i1}, ..., x_{i5}$ as indicator variables for exposure levels;
$Y_i \sim$ binomial (n, π_i) because 'infection' is a binary outcome (but the plausibility of the assumption of independence of infection for mice depends on the experimental conditions);
$g(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$ with the β_k's subject to a corner point or sum-to-zero constraint.

(c) Response: Y_i = number of trips per week;
Explanatory variables: x_{i1} = number of people in the household, x_{i2} = household income, x_{i3} = distance to supermarket;
$Y_i \sim$ Poisson (λ_i) is a simple model for count data with
$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$.

3.2 $a(y) = y$, $b(\theta) = -\theta$, $c(\theta) = \phi \log \theta - \log \Gamma(\phi)$ and $d(y) = (\phi - 1) \log y$.
Hence $\mathrm{E}(Y) = \phi / \theta$ and $\mathrm{var}(Y) = \phi / \theta^2$.

3.3 (a) $\exp\left[\log\theta - (\theta+1)\log y\right]$

(b) $\exp\left(\log\theta - y\theta\right)$

(c) $\exp\left[\log\binom{y+r-1}{r-1} + r\log\theta + y\log(1-\theta)\right]$

3.5 If log (death rate) is plotted against log (age), where the variable age has values 30, …, 65, then the points are close to a straight line.
Simple linear regression produces the estimate

$$\hat{y} = \log(\text{death rate}) = -18.909 + 6.152 \log(\text{age})$$

(with $R^2 = 0.969$ – see Section 6.3.2). This provides a good approximate model although it is based on the Normal distribution not the Poisson distribution.

Estimates of numbers of deaths in each age group can be obtained from $\hat{d}_i = \exp(\hat{y}_i) \times n_i / 100{,}000$. The resulting values are shown in the following table

Age group	Actual deaths	Estimated deaths
30–34	1	1.33
35–39	5	3.02
40–44	5	7.07
45–49	12	11.89
50–54	25	18.73
55–59	38	30.12
60–64	54	57.09
65–69	65	86.68

3.6 (a) $\exp\left\{y_i\left[\log\pi_i - \log(1-\pi_i)\right] + \log(1-\pi_i)\right\}$

(f) As the dose, *x*, increases the probability of death, π, increases from near zero to an asymptotic value of 1.

3.7 Yes, $f(y;\,\theta) = \exp\left(\frac{y}{\phi} - e^{y/\phi}\, e^{-\theta/\phi} - \log\phi - \frac{\theta}{\phi}\right)$.

3.8 The Pareto distribution belongs to the exponential family (see Exercise 3.3(a)), but it does not have the canonical form so this is not a generalized linear model.

3.9 The Normal distribution is a member of the exponential family and has the canonical form. If the exponential function is taken as the link function, then

$$\begin{aligned}\exp(\mu_i) &= \exp\left[\beta_0 + \log(\beta_1 + \beta_2 x_i)\right] \\ &= \beta_0^* (\beta_1 + \beta_2 x_i) \\ &= \beta_1^* + \beta_2^* x_i\end{aligned}$$

which is the linear component.

3.10 $a(y) = \log y$, $b(\theta) = -\theta$, $c(\theta) = \log\theta$ and $d(y) = -\log y$ so that

$$\mathrm{E}\left[a(Y)\right] = -\frac{c'(\theta)}{b'(\theta)} = \frac{1}{\theta}.$$

$U = \dfrac{1}{\theta} - \log y$ so $\mathrm{E}(U) = \dfrac{1}{\theta} - E[\log Y] = 0$

$\mathsf{I} = \mathrm{var}(U) = \theta^{-2}$ by equation (3.15).

CHAPTER 4

4.1 (b) The plot of $\log y_i$ against $\log i$ is approximately linear with a non-zero intercept.

(c) From (4.23) $w_{ii} = \exp\left(\mathbf{x}_i{}^T \boldsymbol{\beta}\right)$.

From (4.24) $z_i = \mathbf{x}_i{}^T \boldsymbol{\beta} + \left[y_i / \exp\left(\mathbf{x}_i{}^T \boldsymbol{\beta}\right)\right] - 1$.

Starting with $b_1^{(0)} = b_2^{(0)} = 1$, subsequent approximations are (0.652, 1.652), (0.842, 1.430), (0.985, 1.334), …, (0.996, 1.327).

(d) The Poisson regression model is
$\log \hat{\lambda}_i = 0.996 + 1.327 \log i$.

4.2 Note: the rows in Table 4.6 on page 68 are wrongly labeled.

(a) y decreases, approximately exponentially, as *x* increases.

(b) log

(c) There is a mistake here; $\mathrm{E}(Y) = 1/\theta$, and $\mathrm{var}(Y) = 1/\theta^2$

Fitted model is $\log(\hat{y}) = 8.4775 - 1.1093x$. .

(e) The model fits the data well; the residuals are small, except for the last observation $(y = 65, x = 5)$ which has $r = 2.467$.

4.3 log-likelihood, $l = -N \log\left(\sigma\sqrt{2\pi}\right) - \dfrac{1}{2\sigma^2}\Sigma(y_i - \log\beta)^2$.

Solve $\dfrac{dl}{d\beta} = \dfrac{1}{\sigma^2\beta}\Sigma(y_i - \log\beta) = 0$ to obtain $\log\hat{\beta} = \bar{y}$.

CHAPTER 5

5.1 $\mathcal{I} = n/\pi(1-\pi)$ so for (a) and (b) Wald statistic $= \dfrac{(y-n\pi)^2}{n\pi(1-\pi)} =$ score statistic.

(c) deviance $= 2\left[y\log\dfrac{\hat{\pi}}{\pi} + (n-y)\log\dfrac{(1-\hat{\pi})}{(1-\pi)} \right]$

where $\hat{\pi} = y/n$.

(d) The 95th percentile of the $\chi^2(1)$ distribution is 3.84 which can be used as the critical value.

(i) Wald/score statistic = 4.44, log-likelihood statistic = 3.07; so the first would suggest rejecting $\pi = 0.1$ and the second would not.

(ii) Both statistics equal zero and would not suggest rejecting $\pi = 0.3$.

(iii) Wald/score statistic = 1.60, log-likelihood statistic = 1.65, so neither would suggest rejecting $\pi = 0.5$.

5.2 Deviance $= 2\sum\left[\dfrac{y_i}{\hat{y}_i} - \log\left(\dfrac{y_i}{\hat{y}_i}\right) - 1 \right]$.

5.3 (a) $\hat{\theta} = \dfrac{N}{\sum \log y_i}$.

(b) $\mathcal{I} = \dfrac{N}{\theta^2}$ so the Wald statistic is $\dfrac{(\theta-\hat{\theta})^2 N}{\theta^2}$.

(c) Approximate 95% confidence limits are given by $\dfrac{(\theta-\hat{\theta})}{\theta}\sqrt{N} = \pm 1.96$,

hence the limits are $\hat{\theta}\Big/\left(1 \mp \dfrac{1.96}{\sqrt{N}}\right)$.

(d) About 1 in 20 intervals should not contain θ.

5.4 (a) (– 1.92, – 0.30).

(b) deviance difference = 26.282 – 19.457 = 6.825; comparing this with $\chi^2(1)$ gives p-value = 0.009 which provides strong evidence that the initial white blood cell count is a statistically significant predictor of survival time.

CHAPTER 6

6.1 (a) Annual per capita consumption of refined sugar decreased by – 4.88 kg per decade (95% confidence interval: – 6.24, – 3.53) using the 97.5th percentile of $t(4)$. Consumption of sugar in manufactured foods increased by 3.62 kg per decade (95% confidence interval: 2.26, 4.98).

(b) Annual per capita total consumption of sugar declined by –1.27 kg per decade (95% confidence interval: – 3.44, 0.91) so we cannot reject the hypothesis of no change ($t = -1.61$, d.f. = 4, p-value = 0.182).

6.2 The quadratic model yield = $2471 + 73.8K - 0.515K^2$ fits the data reasonably well.

6.3

Model	Terms	D	Degrees of Freedom	ΔD
6.6	Age + weight + protein	567.66	16	
6.7	Weight + protein	606.02	17	38.36
(a)	Age + protein	833.57	17	
(b)	Protein	858.65	18	25.08

Using models (6.6) and (6.7) $f = \frac{38.36}{1} / \frac{567.66}{16} = 1.08$

Using models (a) and (b) $f = \frac{25.08}{1} / \frac{833.57}{17} = 0.51$.

In this case neither comparison provides evidence against the null hypothesis that response is unrelated to age. More importantly, however, this example shows that analyses to examine the effect of any variable on the response depend on which other explanatory variables are included in the model (unless the variables are orthogonal).

6.4 (c)

Model	D	Degrees of freedom
Age + BMI	26.571	27
Age	31.636	28

To test the effect of body mass index (BMI), after adjustment for age, use

$$f = \frac{31.636 - 26.571}{28 - 27} / \frac{26.571}{27} = 5.147$$

which is significant compared with the $F(1, 27)$ distribution. So these data suggest that cholesterol level is positively associated with body mass index.

6.5 (a)

Source of variation	Degrees of freedom	Sum of squares	Mean square	f	p-value
Mean	1	350.919			
Between groups	2	7.808	3.904	11.65	<0.001
Residual	28	9.383	0.335		
Total	31	368.110			

Compared with the F (2, 28) distribution the value of f = 11.65 is very significant so we conclude the group means are not all equal. Further analyses are needed to find which means differ.

(b) Using the pooled standard deviation s = 0.5789 (from all groups) and the 97.5th percentage point for t (28), the 95% confidence interval is given by

$$(3.9455 - 3.4375) \pm 2.048 \times 0.5789 \sqrt{\frac{1}{11} + \frac{1}{8}}$$, i.e. (–0.043, 1.059)

6.6

Source of variation	Degrees of freedom	Sum of squares	Mean square	f	p-value
Mean	1	51122.50			
Between workers	3	54.62	18.21	14.45	<0.001
Between days	1	6.08	6.08	4.83	<0.05
Interaction	3	2.96	0.99	0.79	
Residual	32	40.20	1.26		
Total	40	51226.36			

There are significant differences between workers and between days but no evidence of interaction effects.

6.7

Model	Deviance	Degrees of freedom
$\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$	5.00	4
$\mu + \alpha_j + \beta_k$	6.07	6
$\mu + \alpha_j$	8.75	7
$\mu + \beta_k$	24.33	8
μ	26.00	9

(a) $f = \frac{6.07 - 5}{2} \Big/ \frac{5}{4} = 0.43$ so there is no evidence of interaction;

(b) (i) ΔD = 18.26;
(ii) ΔD = 17.25. The data are unbalanced so the model effects are not orthogonal.

6.8

Model	Deviance	Degrees of freedom
$\mu_j + \alpha_j x$	9.36	15
$\mu_j + \alpha x$	10.30	17
$\mu + \alpha x$	27.23	19
μ_j	26.86	18
μ	63.81	20

(a) $f = \dfrac{63.81 - 26.86}{2} \Big/ \dfrac{26.86}{18} = 12.38$ which indicates that the treatment effects are significantly different, if the initial aptitude is ignored.

(b) $f = \dfrac{10.30 - 9.63}{2} \Big/ \dfrac{9.63}{15} = 0.52$ so there is no evidence that initial aptitude has different effects for different treatment groups.

CHAPTER 7

7.1 If dose is defined by the lower end of each dose interval (that is, 0, 1, 10, …, 200), a good model is given by $\text{logit}(\hat{\pi}) = -3.489 + 0.0144\ \text{dose}$.

The Hosmer Lemeshow test of goodness of fit can be obtained from the following table of observed and expected frequencies (the expected frequencies are shown in brackets)

dose	leukemia	other cancer
0	13 (11.6)	378 (379.4)
1 – 49	10 (11.5)	351 (349.5)
50+	25 (24.9)	111 (111.1)

This gives $X^2_{HL} = 0.374$, d. f. = 1, p-value = 0.54 indicating a good fit.

7.2 (a) $\phi = \exp(\beta_1 - \beta_2) = 1$ if and only if $\beta_1 = \beta_2$.

(b) $\phi_j = \exp\left[(\alpha_1 - \alpha_2) + x_j(\beta_1 - \beta_2)\right]$ is constant if $\beta_1 = \beta_2$

7.3 Overall the percentage of women who survived 50 years after graduation (84%) was higher than the percentage of men who survived (67%).

(a) No evidence of differences between years of graduation.

(b) and (c) Higher proportions of science graduates of either sex survived than graduates of other faculties.

(d) The effect seems more pronounced for men (ratio of proportions = 1.40) than for women (ratio = 1.18) but this is not statistically significant.

7.4 (a) $D_0 - D_1 = 2\left[l(b_{\max}) - l(b_{\min})\right] - 2\left[l(b_{\max}) - l(b)\right] = C$

(b) For this hypothesis $D_0 \sim \chi^2(N-1), D_1 \sim \chi^2(N-p)$ so $C \sim \chi^2(p-1)$.

CHAPTER 8

8.2 (a) Satisfaction was lower in apartments or houses than in tower blocks. Contact was lower in tower blocks. However, satisfaction was high in tower blocks with high contact.

(b) A nominal logistic regression model, without interaction terms for house type and contact, fits the data fairly well (Pearson $X^2 = 6.932$, d.f. = 4, p-value = 0.140; deviance = 6.893, d.f. = 4, p-value = 0.142). Most of the parameter estimates are significantly different from zero. The improvement in fit obtained by including interaction terms is not statistically significant.

(c) As satisfaction is an ordinal variable an ordinal logistic regression model is plausible. A proportional odds model without interaction terms for house type and contact fits fairly well (Pearson $X^2 = 11.64$, d.f. = 7, p-value = 0.113; deviance = 11.70, d.f. = 7, p-value = 0.111). However adding interaction terms produces a marginally significant improvement (deviance difference = 6.20, d.f. = 2, p-value = 0,045).

(d) Nominal and ordinal logistic regression models produce similar parameter estimates and similar fitted values in this case. On the grounds of parsimony the ordinal model would be preferred but if interaction terms are included then there is only one less parameter and the interpretation is more complicated. The fit of the nominal logistic regression model is shown in the table below.

house type	contact	satisfaction	observed frequency	estimated frequency	Pearson residual
tower block	low	low	65	59.995	0.646
tower block	high	low	34	39.005	-0.801
tower block	low	medium	54	53.893	0.015
tower block	high	medium	47	47.107	-0.016
tower block	low	high	100	105.112	-0.499
tower block	high	high	100	94.888	0.525
apartment	low	low	130	125.771	0.377
apartment	high	low	141	145.229	-0.351
apartment	low	medium	76	75.222	0.090
apartment	high	medium	116	116.778	-0.072
apartment	low	high	111	116.006	-0.465
apartment	high	high	191	185.994	0.367
house	low	low	67	76.234	-1.058
house	high	low	130	120.766	0.840
house	low	medium	48	48.885	-0.127
house	high	medium	105	104.115	0.087
house	low	high	62	51.881	1.405
house	high	high	104	114.119	-0.947

8.3 Exploratory analysis suggests that there is a difference between treatment groups for the response category 'progressive disease' but not for the other response categories.

(a) Proportional odds model using 'no change' as the reference category gives the following odds ratio estimates: treatment, alternating vs. sequential OR = 1.16 (95% CI: 0.77, 1.75); sex, female vs. male OR = 1.57 (95% CI: 0.88, 2.80).

(b) The model fits poorly: Pearson X^2 = 13.795, d.f. = 7, p-value = 0.055. The table below shows a large residual for 'progressive disease' in females who received 'alternating' treatment.

treatment	sex	stage	observed frequency	expected frequency	Pearson residual
sequential	male	progressive disease	28	36.475	–1.403
sequential	male	no change	45	41.758	0.502
sequential	male	partial remission	29	26.293	0.528
sequential	male	complete remission	26	23.474	0.521
sequential	female	progressive disease	4	6.436	–0.960
sequential	female	no change	12	9.929	0.657
sequential	female	partial remission	5	3.756	0.642
sequential	female	complete remission	2	2.880	–0.519
alternating	male	progressive disease	41	35.772	0.874
alternating	male	no change	44	44.924	–0.138
alternating	male	partial remission	20	24.010	–0.818
alternating	male	complete remission	20	20.294	–0.065
alternating	female	progressive disease	12	6.271	2.288
alternating	female	no change	7	10.767	–1.148
alternating	female	partial remission	3	3.433	–0.234
alternating	female	complete remission	1	2.529	–0.962

(c) For the model in (a) the Wald statistic for alternating compared to sequential treatment is 0.1473/0.2094 = 0.70 (p-value = 0.482) which provides no support for a treatment difference.

(d) Proportional odds models with and without terms for treatment give deviance = 2(–398.4509 + 398.6975) = 0.493, p-value = 0.482 (compared with the $\chi^2(1)$ distribution) – the same result as (c).

(e) Adjacent category and continuation ratio models can be fitted using SAS. A continuation ratio model does not describe the data any better than a proportional odds model. However, the adjacent categories model is better able to describe the treatment difference in the 'progressive disease' category. The choice of link function makes little difference to the results.

8.4 If the probability density function in Figure 8.2 is the Normal distribution then $\pi_i = \Phi(\mathbf{x}_i^T\boldsymbol{\beta})$ from Section 7.3 so the probit model $\Phi^{-1}(\pi_i) = \mathbf{x}_i^T\boldsymbol{\beta}$ is appropriate.

CHAPTER 9

9.2 (a) Claim rates appear to increase with *CAR*, decrease with *AGE* and are higher for *DIST* = 1.

(c) This model is simpler than (b), fits well (deviance = 53.11, d.f. = 60, p-value = 0.72) and gives coefficients (standard errors): *AGE*, – 0.177 (0.018); *CAR*, 0.198 (0.021); *DIST*, 0.210 (0.059), consistent with (a).

9.3 (a) Usual chi-squared test gives $X^2 = 17.65$, d.f. = 2, p-value < 0.001. The same goodness of fit statistic is obtained for the log-linear model with terms for treatment and response categories.

(a) Fitted values are the 'expected frequencies' for a conventional chi-squared test. $X^2 = 17.65$, D = 18.64 with the largest residuals for 'small' response.

(b) For the placebo group the estimated probabilities for the 'small', 'moderate' and 'large' responses are $\hat{\pi}_{11} = 0.638$, $\hat{\pi}_{12} = 0.282$ and $\hat{\pi}_{13} = 0.080$, respectively. For the vaccine group there is a shift of – 1.8373 in the values of $\log\left[\hat{\pi}_{21}/\left(\hat{\pi}_{22}+\hat{\pi}_{23}\right)\right]$ and $\log\left[\left(\hat{\pi}_{21}+\hat{\pi}_{22}\right)/\hat{\pi}_{23}\right]$ to give $\hat{\pi}_{21} = 0.220$, $\hat{\pi}_{22} = 0.426$ and $\hat{\pi}_{23} = 0.354$.

9.5 The log-linear model with all 3 two-way interaction terms produces the same results as the nominal regression model – see solutions for Exercise 8.2 (d).

9.6 (c) The binary logistic regression model with case or control status as the response and ulcer type and aspirin use as the predictor variables produces the same results as the log-linear model with terms *GD* + *CC* + *AP* + *GD*×*CC* + *GD*×*AP* + *CC*×*AP* (see Tables 9.11 and 9.12).

CHAPTER 10

10.1(b) Plots suggest that either the Weibull or exponential distributions with the proportional hazards model may be appropriate, except for a small number of possible outliers.

(c) The estimated shape parameter for the Weibull distribution is $\hat{\lambda} = 0.961$ (95% CI: 0.731, 1.262) suggesting that the simpler exponential distribution could be used.

(d) Two subjects with AG positive, white blood cell count = 100 and survival time = 1 have large residuals, but otherwise the Cox-Snell residuals are consistent with the exponential distribution with a parameter of one. These two points also have the largest deviance residuals.

(e) Survival times decrease with increasing white blood cell count (estimated hazard ratio 0.74; 95% CI: 0.58, 0.94) and were lower for AG negative (estimated hazard ratio 0.36; 95% CI: 0.18, 0.74).

10.2 (a) $S(y) = \dfrac{1}{1+e^{\theta}y^{\lambda}}$, $h(y) = \dfrac{e^{\theta}\lambda y^{\lambda-1}}{1+e^{\theta}y^{\lambda}}$ and $H(y) = \log(1+e^{\theta}y^{\lambda})$.

10.5(a) Exponential, $O(y) = \left[\exp(\theta y) - 1\right]^{-1}$; Weibull, $O(y) = \left[\exp\left(\theta y^{\lambda}\right) - 1\right]^{-1}$; log-logistic, $O(y) = e^{-\theta}y^{-\lambda}$.

10.6 There was an error in the data set on the web – this has now been corrected.

(a) Kaplan Meier estimates $\hat{S}$ suggest better survival for the prednisolone group than the group with no treatment.

(b) Plots of $\log \hat{H}$ and $\log[\hat{S}/(1-\hat{S})]$ against $\log y$ suggest that either the Weibull or log-logistic distribution might be used but the proportional hazards assumption may be inappropriate.

(c) Cox-Snell residuals suggest the log-logistic model describes the data better than a Weibull model. The coefficient for prednisolone group vs. on treatment group is 1.33 (95% CI: 0.28, 2.38) is significantly different from zero (p-value = 0.013).

CHAPTER 11

11.1

	Intercept (s.e.)	Slope (s.e.)
(a) Pooled	40.768 (6.617)	0.769 (0.092)
(b) Data reduction		
(i) dogs, ignoring conditions	37.308 (20.480)	0.774 (0.277)
(ii) conditions, ignoring dogs	40.260 (6.451)	0.768 (0.090)
(c) Random effects		
(i) dogs random, conditions fixed	45.393 (8.785)	0.728 (0.108)
(ii) conditions random, dogs fixed	44.863 (7.281)	0.629 (0.126)
(iii) both random	68.430 (13.348)	0.458 (0.161)
(d) GEE		
(i) conditions fixed *	41.531 (7.095)	0.781 (0.093)
(i) dogs fixed*	37.119 (5.576)	0.774 (0.097)

* robust standard errors

Results from the model with both 'dog' and 'condition' random effects are clearly inconsistent with estimates from the other models. Effects of 'condition' are smaller than 'dog' effects and can be treated as fixed without affecting the estimates greatly. The best estimates are probably from the random effects or GEE models with 'condition' as a fixed effect, 'dog' as a random effect and robust estimates of standard errors.

11.3 (a) For the youngest age group neither treatment appeared to work well. For the oldest age group most children recovered fully regardless of treatment.

(b) If the clustering is not taken into account there are significant treatment and age effects. If clustering is taken into account the odds ratio for treatment is 0.375 (95% CI: 0.113, 1.245) but the model cannot describe the age effects.

(c) Nominal logistic regression fits well (Pearson $X^2 = 4.56$, d.f. = 4, p-value = 0.335). There is no significant difference between the treatment groups. This model does detect a significant difference between the age groups <2 and ≥ 6 years but the confidence interval is wide due to few children aged ≥ 6.

CHAPTER 12

12.1 (a) H_1: Infection is endemic ($\theta > 0.5$).

	Observed data	
Prior	5 out of 10 positive	1 out of 10 positive
$P(H_1)=0.5$	0.408	0.002
$P(H_1)=0.99$	0.986	0.198

(b)

	Observed data	
Prior	5 out of 10 positive	1 out of 10 positive
$P(H_1)=0.5$	0.491	0.005
$P(H_1)=0.99$	0.990	0.352

You can use the "Bolstad" library in R to do the calculations. For example, to calculate the top-left cell for part (a)

```
> library(Bolstad)
> theta.space<-(0:10)/10
> theta.mass<-c(rep(0.5/6,6),rep(0.5/5,5))
> sum(theta.mass) # Check it sums to one
> sum(theta .mass[7:11]) # Prior probability for H1
> post<-binodp(5,10,pi=theta.space,pi.prior=theta.mass,ret=TRUE)
> sum(post$posterior[7:11])
```

Thinking about the top-left cell of the table, there is a lot of posterior probability that is "missed" by the granular prior in part (a) between 0.51 and 0.60. A finer parameter space counts this large area of posterior probability, and hence the increased posterior probability for H_1.

12.3 (a) prior P(LHR>0) = 0.950

(b) posterior P(LHR>0) = 0.999

Using R:

```
> post<-function(mu_0,mu_l,sigma_0,sigma_l){
> mu_p<-((mu_0*(sigma_l^2))+(mu_l*(sigma_0^2)))/((sigma_l^2)+(sigma_0^2))
> var_p<-((sigma_l^2)*(sigma_0^2))/((sigma_l^2)+(sigma_0^2))
> sigma_p<-sqrt(var_p)
> print(mu_p)
> print(sigma_p)
> }
```

> 1-pnorm(0,0.3137,0.1907) # prior prob trt is effective

> post(mu_0=0.3137,sigma_0=0.1907,mu_l=0.580,sigma_l=0.2206)

> 1-pnorm(0,0.428,0.144) # prior prob trt is effective

12.4 (a) Investigator's prior is 1 overdose in 200 (mean rate = 0.005). Be(2,200) prior. (b) Posterior = Beta (2+0, 91-0+200) or Be(2,291), gives a mean = 2 / 293 = 0.00683 or 1 overdose per 146.5 released prisoners. So the posterior mean has decreased compared to the prior, despite there being no overdoses.

CHAPTER 13

13.1 Results depend on 11 randomly generated values for θ

θ	Hypothesis	P(θ) Prior	P(y|θ) Likelihood	P(y|θ)×P(θ) Likelihood×Prior	P(θ|y) Posterior
0.0008	H_0	0.04	0.0000	0.0000	0.0000
0.1082	H_0	0.04	0.0000	0.0000	0.0000
0.1306	H_0	0.04	0.0001	0.0000	0.0000
0.2121	H_0	0.04	0.0011	0.0000	0.0004
0.3994	H_0	0.04	0.0421	0.0017	0.0157
Sum		0.2000			0.0161
0.5105	H_1	0.1333	0.1272	0.0170	0.1576
0.5180	H_1	0.1333	0.1345	0.0179	0.1666
0.6243	H_1	0.1333	0.2352	0.0314	0.2914
0.8123	H_1	0.1333	0.1852	0.0247	0.2294
0.8899	H_1	0.1333	0.0708	0.0094	0.0877
0.9134	H_1	0.1333	0.0413	0.0055	0.0512
Sum		0.8000		0.1076	0.9839

For this data the posterior probability that the village is endemic is slightly higher 0.9839 (compared to 0.9545 for equally spaced prior parameter space). A restriction would be to have at least one generated value below 0.5 and at least one above 0.5. Similarly you could say at least two, or three, etc. This restriction is equivalent to using a prior distribution that covers the likely values for θ.

13.2 (a) R code:

> theta=vector(1000,mode="numeric")

> accept=vector(1000,mode="numeric")

> theta[1]=0.5

> accept[1]=NA

> for (i in 1:1000){

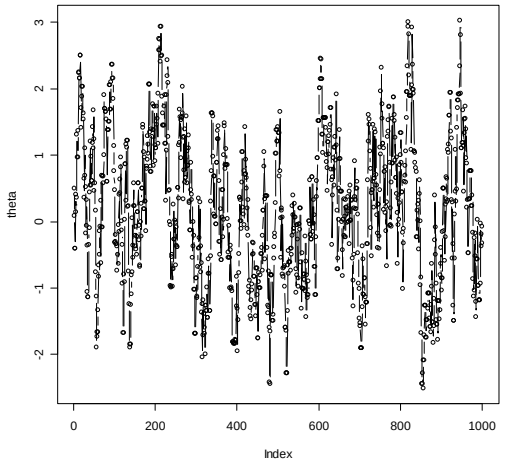
> Q<-runif(1,-1,1) # proposal U[-1,1]

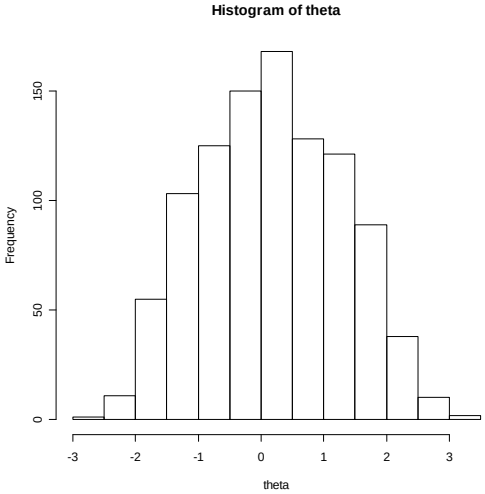
> theta_star<-theta[i]+Q

> pstar<-dnorm(theta_star,mean=0,sd=1)

```
>    p<-dnorm(theta[i],mean=0,sd=1)
>    alpha<-min(pstar/p,1)
>    U<-runif(1,0,1) # acceptance r.v.
>    if (U>=alpha){theta[i+1]=theta[i]
>    accept[i]=0}
>    if (U<alpha){theta[i+1]=theta_star
>    accept[i]=1}
> }
> sum(accept[2:1000])/999 # acceptance probability
> hist(theta)
> plot(theta,type='b')
```

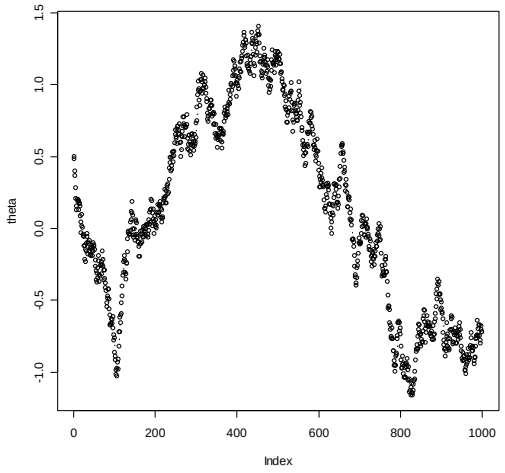
Acceptance rate = 0.799

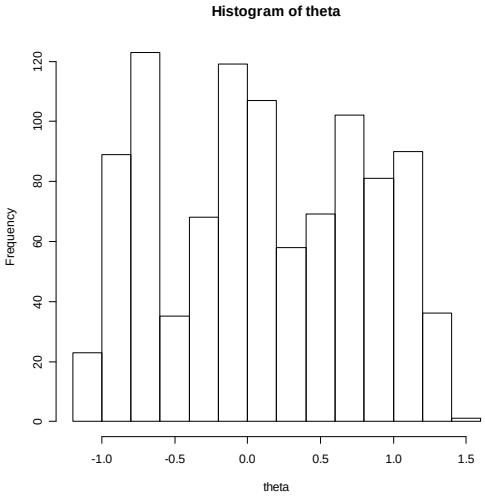




(b) For Q~U[-0.1,0.1]

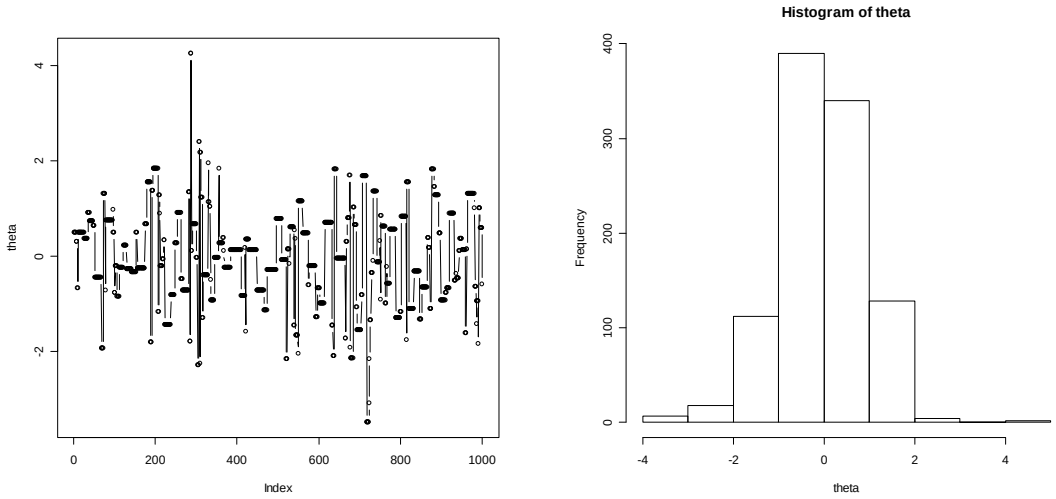
Acceptance rate = 0.974





For Q~U[-10,10]

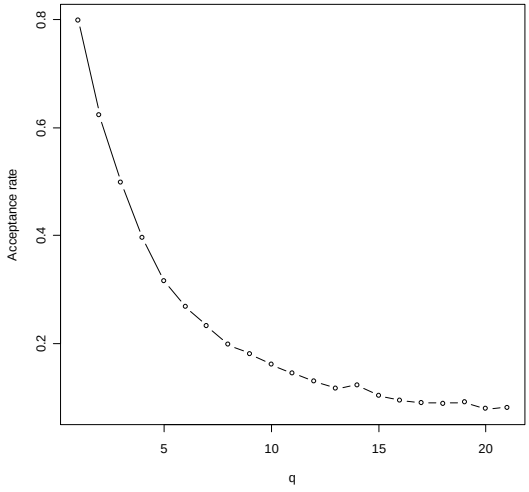
Acceptance rate = 0.161



The smaller Uniform proposal takes only small steps (which are often accepted) and barely reaches the tails of the standard Normal distribution. In contrast the larger Uniform proposal has reached the tails, but the history also shows periods where the proposal is not accepted. This means certain modes will be over-represented, although the histogram above has smoothed any multi-modality.

(c) $q \approx 2.17$ gives an approximately 60% acceptance rate

(d) based on 10,000 samples

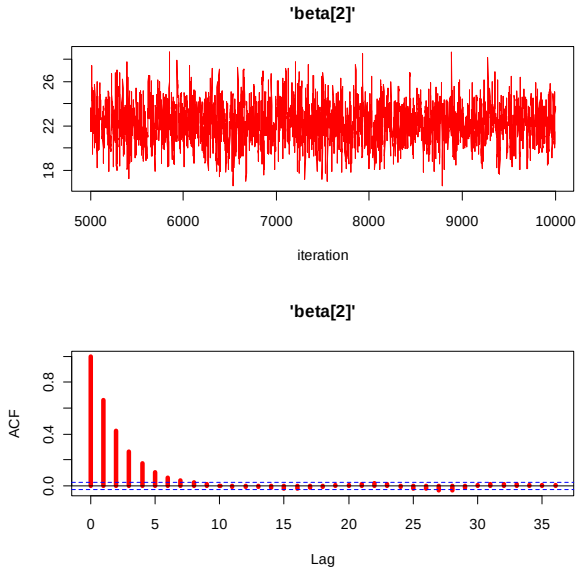


13.3 (c) monotonically increasing (S-shaped), skewed towards larger probabilities (smallest probability is around 10%)

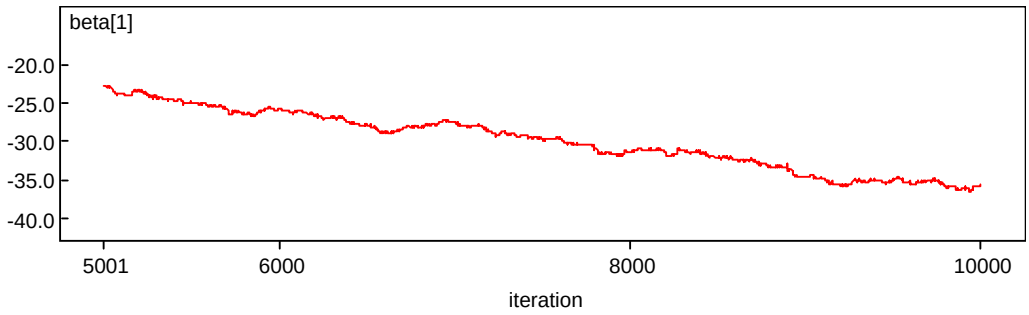
(d) Assumptions: Independence between observations, vague normal priors for unknowns, Binomial distribution for success probability

(e) A constant risk of p=0.63

(f) The results of this exercise change greatly depending on whether you used the Metropolis-Hastings or Gibbs sampler. The default in BRugs (version 0.4-1) is the Gibbs sampler which gives results like those below.



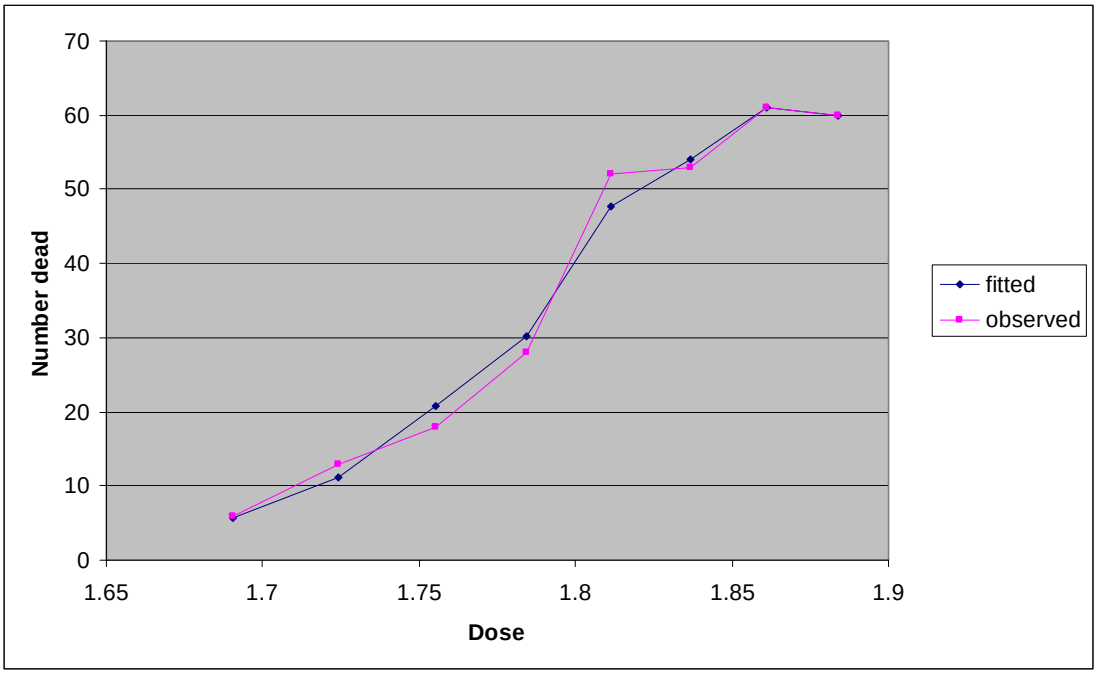
The default in WinBUGS (version 1.4) is the Metropolis-Hastings, which gives much poorer chains as shown below.



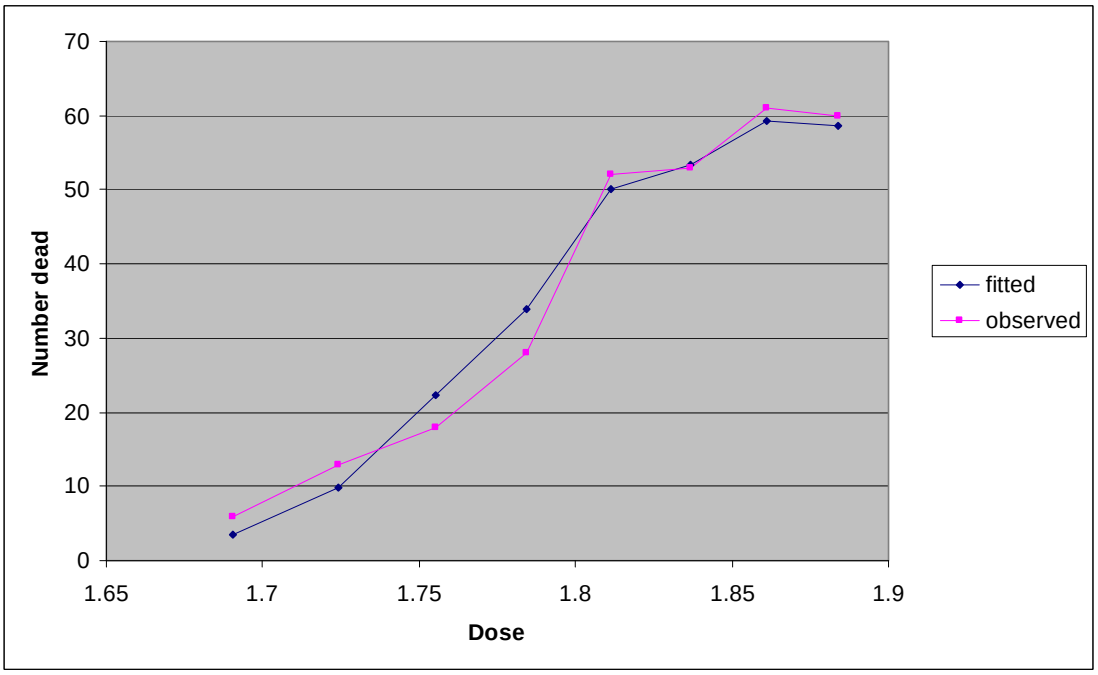
(g) The chains based on a centred dose using M-H sampling will improve dramatically due to the likelihood becoming less precipitous (more like a hill than a sharp ridge). The centred chains using Gibbs sampling don't change much.

(h) Slow decline in the deviance (with M-H sampling) matched by a slow change in the beta parameters. The estimates are slowly creeping towards a "good" solution.

(i)



(j)



The extreme value link clearly gives a better fit to the observed data.

13.4 (a) Estimate $D(\mathbf{y} \mid \overline{\boldsymbol{\beta}})$ using the sampled value of the deviance (out of all 1000 samples) that is closest to $\overline{\boldsymbol{\beta}}$ (i.e. the pair of β_1 and β_2 that are closest to the mean).

(b)–(d)

Version	$\overline{D(\mathbf{y} \mid \boldsymbol{\beta})}$ (Dbar)	$D(\mathbf{y} \mid \hat{\boldsymbol{\beta}})$ (Dhat)	p_D	DIC
dicStats()	31.266	29.649	1.617	32.883

Mean of β for $\hat{\beta}$	31.266	29.65	1.616	32.882
Median of β for $\hat{\beta}$	31.266	29.66	1.606	32.872
Half variance of $D(.)$	31.266		1.505	32.771

There's very little difference from using the mean or median in this example, this is because the three-dimensional deviance was quite symmetrical (spherical). The half-variance method gives an almost identical result too, it might have been different if the deviance distribution was more skewed.

(e)

Version	$\overline{D(\mathbf{y} \mid \boldsymbol{\beta})}$ (Dbar)	$D(\mathbf{y} \mid \hat{\boldsymbol{\beta}})$ (Dhat)	p_D	DIC
dicStats()	39.473	37.443	2.030	41.503
Mean of β for $\hat{\beta}$	39.473	37.45	2.023	41.496
Median of β for $\hat{\beta}$	39.473	37.64	1.833	41.306
Half variance of $D(.)$	39.473		1.888	41.361

The extreme value link is always a better model, no matter how $D(\mathbf{y} \mid \hat{\boldsymbol{\beta}})$ is calculated. The histories for the extreme value link showed poor mixing, and a fairer comparison would have been achieved using a longer burn-in and thinning for both link functions.

(f) Multiple posterior means or medians (use half the variance of $D(.)$ in this case). A non-spherical deviance (thinking only two parameters) could have a posterior mean far from the posterior variance, which may then give different DIC estimates depending on which method was used.

CHAPTER 14

14.3 (a) $p_D = 2.90$, DIC = 219 (5,000 burn-in and 10,000 sample)

(b) $p_D = 1.94$, DIC = 221 (5,000 burn-in and 10,000 sample)

So really very little difference between the models, so go for Exponential on the grounds of parsimony (exactly as per section 10.7).

14.4 (a) r = –0.30

(b) r = –0.35, 95% PI = –0.68, 0.06. It might seem surprising that the correlation is negative. The reason is because those with the lowest intercepts had the most amount of room for improvement (intercepts lower than the average paired with slopes higher

than the average). Conversely those who were already scoring close to 100 had little room for improvement (intercepts higher than the average paired with slopes lower than the average).

14.5

(b)

```
model{
# likelihood
        for(subject in 1:N) { # loop in subject
        ability[subject,1:T] ~ dmnorm(mu[subject,1:T],omega.obs[1:T,1:T]);
        for(time in 1:T) { # loop in time
                mu[subject,time] <- alpha.c[group[subject]] +
(beta[group[subject]]*time);
        } # end of time loop
        } # end of subject loop
# inverse variance-covariance matrix
    omega.obs[1,1] <- tau.obs; omega.obs[T,T] <- tau.obs;
    for (j in 2:T-1){ omega.obs[j, j] <- tau.obs*(1+pow(rho,2));} # diagonal
    for (j in 1:T-1){ omega.obs[j, j+1] <- -tau.obs*rho;
                    omega.obs[j+1, j] <- omega.obs[j, j+1];} # symmetry
    for (i in 1:T-1) {
      for (j in 2+i:T) {
        omega.obs[i, j] <- 0; omega.obs[j, i] <- 0;
      }
    }
# priors
    tau.obs ~ dgamma(0.001,0.001);
    rho~dunif(-0.99,0.99); # correlation parameter
    beta[1] ~ dnorm(0,1.0E-4); # Linear effect of time (group=A)
    beta[2] ~ dnorm(0,1.0E-4);
    beta[3] ~ dnorm(0,1.0E-4);
    alpha.c[1]~dnorm(0,1.0E-4); # Centred intercept (group=A)
    alpha.c[2]~dnorm(0,1.0E-4);
    alpha.c[3]~dnorm(0,1.0E-4);
# scalars
```

```
    b.diff[1]<-beta[2]-beta[1];
    b.diff[2]<-beta[3]-beta[1];
    alpha[1]<-alpha.c[1]+50; # re-adjust intercept
    alpha[2]<-alpha.c[2]+50;
    alpha[3]<-alpha.c[3]+50;
    a.diff[1]<-alpha[2]-alpha[1];
    a.diff[2]<-alpha[3]-alpha[1];
      var.obs <- 1 / (tau.obs*(1-pow(rho,2))); # sigma^2
}
```

(d) roughly means that we have observed the covariance matrix in a previous sample of 500, and are confident that it is applicable to this data

(e) Both posteriors estimates of $\hat{\mathbf{V}}$ show a steadily increasing variance over time (values along the diagonal). As expected using a strong prior has produced a posterior with very similar values to the prior. Using the strong prior the estimated number of parameters was only just above the number of parameters from the independent correlation model (7.6 compared to 7.0). The small number parameters needed for this unstructured covariance combined with the high degree of freedom in the estimates has given the best fit.