M3/M4
S2

Imperial College
London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2015

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science.

Statistical Modelling II

Date: Wednesday, 27 May 2015. Time: 10.00am – 12.00noon. Time allowed: 2 hours.

This paper has FOUR questions.

Candidates should use ONE main answer book.

Supplementary books may only be used after the main book is full.

Statistical tables will not be provided.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.
- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.
- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

Raw mark	up to 12	13	14	15	16	17	18	19	20
Extra credit	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4

- Each question carries equal weight.
- Calculators may not be used.

1. a) Recall that the logit link function is

$$g(u) = \log\left(\frac{u}{1-u}\right), \quad u \in (0,1).$$

Find the inverse of the logit function.

b) Consider a Binomial generalized linear model (GLM) with 3 covariates and an intercept term. Suppose that the number of trials for each binomial observation is equal to one, i.e. $Y_i \sim \text{Binomial}(1, \pi_i)$. Using data, consisting of 100 observations, Statistician A fits this model with the logit link function. The estimated parameters using these data are:

$$\widehat{\beta}_0 = 5, \quad \widehat{\beta}_1 = 1, \quad \widehat{\beta}_2 = 2, \quad \widehat{\beta}_3 = -1.$$

Predict the response for an observation with the covariates:

$$x_1 = \frac{1}{2}, \quad x_2 = \frac{1}{3}, \quad x_3 = \frac{1}{4}$$

c) Statistician B fits the same Binomial GLM as Statistician A but without an intercept term. Comment on the difference of the deviances of the models fitted by Statistician A and B. State which statistical test should be used to compare the two models.

d) Consider the independent random variables $Y_1, \ldots, Y_n$ where $\mathrm{E}(Y_i) = \mu_i$ and $\text{var}(Y_i) = \phi V(\mu_i)$. Recall that the quasi-deviance is defined as $D = \sum_{i=1}^n D_i$ where

$$D_i = -2\phi Q_i \quad \text{and} \quad Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt.$$

By selecting $V(\cdot)$ appropriately, derive the quasi-deviance for the case where $Y_i \sim \text{Binomial}(m_i, \pi_i)$ with $\mathrm{E}(Y_i) = m_i \pi_i$ and $\text{var}(Y_i) = m_i \pi_i (1 - \pi_i)$.

e) Statistician C believes both Statistician A and B are incorrect and proceeds to fit exactly the same models, but using a quasi-Binomial response. State which statistical test should be used to compare a pair of nested quasi-Binomial models?

2. a) Consider the following random effects model:

$$Y_{ij} = \mu + \nu_j + \epsilon_{ij}, \quad i = 1, \ldots, K, j = 1, \ldots, m,$$

where the random effect, $\nu_j \sim N(0, \sigma_\nu^2)$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ and μ is the fixed effect. The ν_j and ϵ_{ij} are all independent. Further, consider the following statistics:

$$SST = \sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij} - \overline{Y})^2, \quad SSA = \sum_{i=1}^{K}\sum_{j=1}^{m}(\overline{Y}_{\bullet j} - \overline{Y})^2, \quad SSE = \sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij} - \overline{Y}_{\bullet j})^2$$

where

$$\overline{Y}_{\bullet j} = \frac{1}{K}\sum_{i=1}^{K} Y_{ij} \quad \text{and} \quad \overline{Y} = \frac{1}{mK}\sum_{i=1}^{K}\sum_{j=1}^{m} Y_{ij}.$$

Prove $SST = SSA + SSE$

b) Show that the correlation between observations in the same group is

$$\frac{\sigma_\nu^2}{\sigma_\epsilon^2 + \sigma_\nu^2}$$

c) Consider the maths school exam dataset from lectures. The dataset consists of students maths scores (score) from 5 schools (school). The statistical package R was used to estimate the model presented in part a), producing the output below.

```
mylme <- lmer(score~1+(1|school),data=mathres,REML=FALSE)

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: score ~ 1 + (1 | school)
   Data: mathres

     AIC      BIC   logLik deviance df.resid
  2119.6   2130.9  -1056.8   2113.5      323

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6773 -0.6110  0.1744  0.7170  1.7248

Random effects:
 Groups   Name        Variance Std.Dev.
 school   (Intercept)  2.058   1.435
 Residual             37.411   6.116
Number of obs: 326, groups:  school, 5

Fixed effects:
            Estimate Std. Error t value
(Intercept)  13.8023     0.7255   19.02
```

Confirm the reported value of the AIC given in the R output using the reported log likelihood.

d) If we take an empirical Bayes point of view and estimate ν_i using its posterior distribution, we can use R to compute:

```
> ranef(mylme)
$school
  (Intercept)
1  -2.0959
2   1.2653
3  -0.7329
4   1.1639
5   0.3995
```

What is the expected maths exam score for a student from school 3?

e) The model was fitted using the maximum likelihood (ML) estimation. An alternative approach is the restricted maximum likelihood (REML) estimation. In one sentence, explain the intuition behind the REML approach. State one desirable or undesirable feature of REML estimation.

3. a) The exponential family of distributions take the general form:

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

Consider the Poisson distribution with probability mass function

$$P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, \ldots, \text{and } \lambda > 0.$$

Show that the Poisson distribution is a member of the exponential family. Further, stating results from lectures, confirm the well-known property for the Poisson distribution that $\mathrm{E}(Y) = \mathrm{var}(Y)$.

b) Given a set of data, a statistician fits a Poisson GLM with 3 covariates and an intercept. The R summary of the fit is given below.

```
Call:
glm(formula = y ~ x1 + x2 + x3, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1443  -1.0359  -0.3049   0.6759   3.9302

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.3567352  0.1579328  -2.259   0.0239 *
x1          -0.0005985  0.0024105  -0.248   0.8039
x2           0.2181490  0.0235115   9.278   <2e-16 ***
x3           3.2483116  0.1523890  21.316   <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: 828.54  on 99  degrees of freedom
Residual deviance: 219.62  on 96  degrees of freedom
AIC: 567.62

Number of Fisher Scoring iterations: 5
```

In the R summary above there is an asterisk $*$ in the `Intercept` coefficient row. State the hypothesis test conducted on the intercept coefficient and carefully explain the result of the test.

c) A commonly used residual is Pearson's residual which is defined as

$$r_p^i = \frac{y_i - \mu_i}{\sqrt{V(\mu_i)}},$$

where we use the theoretical form, involving the expected responses μ_i rather than its fitted value $\widehat{\mu}_i$. Find the expectation and variance of r_p^i for a Poisson GLM.

d) Consider the diagnostic plot, presented in Figure 1, for the model fitted in part b) . After viewing Figure 1, the statistician believes the model suffers from overdispersion. Explain what is meant by overdispersion. Further, using your answer from part c) or otherwise, explain why Figure 1 supports the statisticians claim.

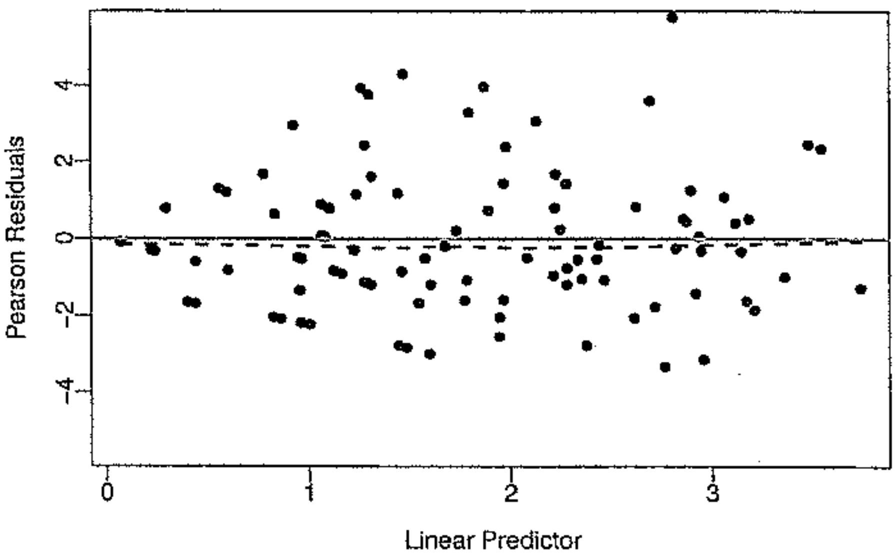

Figure 1: Pearson's residual plotted against the linear predictor for the Poisson GLM fitted in part b), with a smoothed fitted line of the residuals (dashed line).

e) The statistician calculates Pearson's estimate of ϕ as $\widehat{\phi}_P \approx 2.259256$, and proceeds to re-evaluate the model in R as follows:

```
summary(myglm,dispersion=phihat)

Call:
glm(formula = y ~ x1 + x2 + x3, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1443  -1.0359  -0.3049   0.6759   3.9302

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.3567352    ?????     ---     ---
x1          -0.0005985    ?????     ---     ---
x2           0.2181490    ?????     ---     ---
x3           3.2483116    ?????     ---     ---
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 2.259256)

    Null deviance: 828.54  on 99  degrees of freedom
Residual deviance: 219.62  on 96  degrees of freedom
AIC: 567.62

Number of Fisher Scoring iterations: 5
```

The standard errors reported in the R summary have been censored with "?????". How do you expect these values to change from the previous analysis?

4. a) The exponential family of distributions take the general form:

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

Consider the Binomial distribution with probability mass function

$$P(Y = y; m, \pi) = \binom{m}{y}\pi^y(1-\pi)^{m-y},$$

where $m \in \mathbb{Z}$, $0 \leq \pi \leq 1$, $y = 0, 1, \ldots, n$. Express the Binomial distribution in exponential family form.

b) Show that the deviance for a GLM with Binomial responses, i.e. independent $Y_1, \ldots, Y_n$ with $Y_i \sim \text{Binomial}(m_i, \pi_i)$, has the form

$$D = 2\sum_{i=1}^{n}\left\{y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - \widehat{\mu}_i}\right)\right\}$$

c) Consider the case with $m_i = 1$, for $i = 1, \ldots, n$. Show that the deviance in part b) can be written as

$$D = -2\sum_{i=1}^{n}\left\{y_i \log\left(\frac{\widehat{\mu}_i}{1-\widehat{\mu}_i}\right) + \log(1-\widehat{\mu}_i)\right\}$$

d) We shall continue with the $m_i = 1$ case and use the logit link function:

$$\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right),$$

where $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$ is the linear predictor. Compute $\frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta}; \boldsymbol{y})$ and then show that

$$\sum_{j=1}^{p}\beta_j\frac{\partial}{\partial\beta_j}\ell(\boldsymbol{\beta};\boldsymbol{y}) = \sum_{i=1}^{n}\left[(y_i - \mu_i)\log\left(\frac{\mu_i}{1-\mu_i}\right)\right],$$

where $\ell(\boldsymbol{\beta};\boldsymbol{y})$ is the log-likelihood. Then, using the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$, show that the deviance in part c) can be written as

$$D = -2\sum_{i=1}^{n}\left\{\widehat{\mu}_i \log\left(\frac{\widehat{\mu}_i}{1-\widehat{\mu}_i}\right) + \log(1-\widehat{\mu}_i)\right\}$$

e) Using part d) explain why the deviance is not an appropriate measure of the goodness of fit in the case where $m_i = 1$ for $i = 1, \ldots, n$.

Imperial College
London

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

2015

This paper is also taken for the relevant examination for the Associateship.

M3S2/M4S2

Statistical Modelling II (Solutions)

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. a) Simple calculations yields $\boxed{\text{seen} \Downarrow}$

$$z = \log\left(\frac{u}{1-u}\right) \implies e^{-z} = \frac{1}{u} + 1 \implies u = \frac{1}{1+e^{-z}}.$$

Hence the inverse is $\boxed{2}$

$$g^{-1}(u) = \frac{1}{1+e^{-u}}$$

$\boxed{\text{meth seen} \Downarrow}$

b) For the given covariates the linear predictor is

$$\widehat{\beta}_0 + \widehat{\beta}_1 \frac{1}{2} + \widehat{\beta}_2 \frac{1}{3} + \widehat{\beta}_3 \frac{1}{4}$$
$$5 + (1) \cdot \frac{1}{2} + (2) \cdot \frac{1}{3} + (-1) \cdot \frac{1}{4} = \frac{71}{12}.$$

To get the prediction, we use the inverse logit function from part a) $\boxed{3}$

$$\frac{1}{1+e^{-71/12}}$$

$\boxed{2}$

c) The deviance for Statistician B's model will be greater or equal to than the deviance for Statistician A's because Statistician B's has one fewer parameters. $\boxed{\text{seen} \Downarrow}$

$\boxed{1}$

In the Binomial GLM, the dispersion parameter is $\phi = 1$, thus we use the $D_0 - D_1 \sim \chi^2_1$ approximation where D_0 is the deviance of the smaller model (Statistician B's model) and D_1 is the deviance of the larger model. This holds under the null hypothesis which corresponds to the smaller model being true. $\boxed{3}$

d) For the Binomial distribution, we select $V(\mu_i) = \mu_i(m_i - \mu_i)/m_i$ where $\mu_i = \mathrm{E}(Y_i) = m_i \pi_i$. Then $\boxed{\text{sim. seen} \Downarrow}$

$$\begin{aligned}\frac{1}{m_i}\phi Q_i &= \int_{y_i}^{\mu_i} \frac{y_i - t}{t(m_i - t)} dt \\ &= \int_{y_i}^{\mu_i} \frac{y_i/m_i}{t} + \frac{(y_i/m_i) - 1}{m_i - t} dt\end{aligned}$$

$\boxed{2}$

$$\begin{aligned}\implies \phi Q_i &= \int_{y_i}^{\mu_i} \frac{y_i}{t} + \frac{y_i - m_i}{m_i - t} dt \\ &= [y_i \log(t) - (y_i - m_i)\log(m_i - t)]_{y_i}^{\mu_i} \\ &= -y_i \log\left(\frac{y_i}{\mu_i}\right) - (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - \mu_i}\right)\end{aligned}$$

$\boxed{2}$

$$\implies D = \sum_{i=1}^{n} -2\phi Q_i = 2\sum_{i=1}^{n}\left\{y_i \log\left(\frac{y_i}{\mu_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - \mu_i}\right)\right\}$$

Students have seen worked examples for other members of the exponential family. $\boxed{1}$

e) For the quasi-Binomial GLM, the dispersion parameter is modelled as a free parameter; therefore the F test is required. The approximate test is seen $\Downarrow$ 1

$$\frac{(D_0 - D_1)/(p-q)}{D_1/(n-p)} \sim F_{(p-q),(n-p)}.$$

where D_0 is the smaller quasi-Binomial model and D_1 is the other quasi-Binomial model. 2

We recognise that $p = 4$, $q = 3$ and $n = 100$, so the test statistic is compared to the $F_{1,96}$ distribution. 1

2. a) Starting with the definition of SST we have $\boxed{\text{seen} \Downarrow}$ $\boxed{2}$

$$\begin{aligned} SST =& \sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij}-\overline{Y})^2 = \sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij}-\overline{Y}_{\bullet j}+\overline{Y}_{\bullet j}-\overline{Y})^2 \\ =& SSA + SSE + 2\sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij}-\overline{Y}_{\bullet j})(\overline{Y}_{\bullet j}-\overline{Y}). \end{aligned}$$

The rightmost summand is fact zero since

$$\sum_{i=1}^{K}\sum_{j=1}^{m}(Y_{ij}-\overline{Y}_{\bullet j})(\overline{Y}_{\bullet j}-\overline{Y}) = \sum_{j=1}^{m}(\overline{Y}_{\bullet j}-\overline{Y})\sum_{i=1}^{K}(Y_{ij}-\overline{Y}_{\bullet j})$$

and $\boxed{2}$

$$\sum_{i=1}^{K}(Y_{ij}-\overline{Y}_{\bullet j}) = \sum_{i=1}^{K}Y_{ij} - \sum_{i=1}^{K}\overline{Y}_{\bullet j} = K\overline{Y}_{\bullet j} - K\overline{Y}_{\bullet j} = 0$$

$\boxed{2}$

b) The correlation for observations in the same group is:

$$\text{corr}(Y_{1,j}, Y_{2,j}) = \frac{\mathrm{E}\left[(Y_{1,j}-\mathrm{E}(Y_{1,j}))\,(Y_{2,j}-\mathrm{E}(Y_{2,j}))\right]}{\sqrt{\mathrm{var}(Y_{1,j})\,\mathrm{var}(Y_{2,j})}}$$

We have that $\mathrm{var}(Y_{1,j}) = \mathrm{var}(Y_{2,j}) = \sigma_\epsilon^2 + \sigma_\nu^2$, and thus the denominator is $\sigma_\epsilon^2 + \sigma_\nu^2$. Next, we have $Y_{i,j} - \mathrm{E}(Y_{i,j}) = \nu_j + \epsilon_{ij}$, by noting that $\mathrm{E}(Y_{i,j}) = \mu$ and rearranging the model equation. Therefore the numerator is $\boxed{2}$ $\boxed{1}$

$$\begin{aligned} \mathrm{E}\left[(Y_{1,j}-\mathrm{E}(Y_{1,j}))\,(Y_{2,j}-\mathrm{E}(Y_{2,j}))\right] &= \mathrm{E}\left[(\nu_j+\epsilon_{1,j}))\,(\nu_j+\epsilon_{2,j}))\right] \\ &= \mathrm{E}(\nu_j^2) + \mathrm{E}(\nu_j(\epsilon_{1,j}+\epsilon_{2,j})) + \mathrm{E}(\epsilon_{1,j}\epsilon_{2,j}) \\ &= \mathrm{E}(\nu_j^2) = \sigma_\nu^2, \end{aligned}$$

where we used the independence between the ν_j and ϵ_{ij}. Plugging in these results yields the require solution. $\boxed{2}$

c) Recall the definition of AIC: $\boxed{2}$

$$\text{AIC} = -2\text{loglik} + 2\text{number of parameters}$$

The model considered has 1 fixed effect parameter, μ, and the 2 variance components. Therefore,

$$\begin{aligned} \text{AIC} &= (2 \times 1056.8) + (2 \times 3) \\ &= 2119.6 \end{aligned}$$

$\boxed{1}$

d) First the estimated fixed effect is $\widehat{\mu} = 13.8023$ and from the R output we have that the predicted random effect for school 3 is $\widehat{\nu_3} = -0.7329$ adding these together gives the result

$$13.8023 - 0.7329 = 13.0694$$

$\boxed{2}$

e) The intuition behind the REML estimation is to use a transformation to eliminate the fixed effects: More precisely, use an appropriate matrix L such that $L^T X = 0$ such that, for general normal linear mixed models, we have:

$$\underbrace{L^T \boldsymbol{Y}}_{\boldsymbol{B}} = \underbrace{L^T X}_{0} \boldsymbol{\beta} + L^T Z \boldsymbol{\nu} + L^T \boldsymbol{\epsilon}$$
$$\implies \boldsymbol{B} = L^T Z \boldsymbol{\nu} + L^T \boldsymbol{\epsilon},$$

which has no fixed effects — this is for the general normal linear mixed normal introduced in lectures, where X is the design matrix, Z is the model matrix for the random effects, $\boldsymbol{\nu}$ is the random effect and $\boldsymbol{\epsilon}$ is the vector of errors.

Note that the student need not present these equations — a worded answer will suffice. $\boxed{2}$

Two possible answers are:

* REML usually yields less biased estimators than the ML approach.
* A drawback of REML is that the comparison of the fixed effects between two models is not possible as the two restricted likelihoods are not comparable. $\boxed{2}$

3. a) The probability mass function for the Poisson(λ) distribution can be written as: seen ⇓

$$\exp\left\{y\log(\lambda) - \lambda - \log(y!)\right\}.$$ 1

Thus we identify

$$\theta = \log(\lambda), \phi = 1, a(\phi) = 1, b(\theta) = \lambda = \exp(\theta), c(y,\phi) = \log(y!).$$

2

From the results in lectures, we know that $\mu \equiv \mathrm{E}(Y) = b'(\theta)$ and $\mathrm{var}(Y) = a(\phi)b''(\theta)$ and therefore for the Poisson distribution $E(Y) = \lambda = \mathrm{var}(Y)$ 3

b) The null hypothesis being tested is $H_0 : \beta_0 = 0$ where β_0 is the corresponding coefficient for the intercept term. R reports this as significant, meaning there is sufficient evidence to reject the null, at the 5% level — thus suggesting inclusion of the intercept in the model.

3

c) From lectures, we know, for members of the exponential family, that $\mathrm{E}(Y_i) = \mu_i$ and $\mathrm{var}(Y_i) = \phi V(\mu_i)$; in particular for the Poisson distribution $\mu_i = \lambda_i$, $V(\mu_i) = \lambda_i$ and $\phi = 1$ (from Q3a). Therefore unseen ⇓ 2

$$\mathrm{E}(r_p^i) = \frac{1}{\sqrt{V(\mu_i)}}\mathrm{E}(Y_i - \mu_i) = 0$$

and 1

$$\mathrm{var}(r_p^i) = \frac{1}{V(\mu_i)}\mathrm{var}(Y_i) = \phi = 1.$$

1

d) Overdispersion means that the observed variance of the response is larger than the variation implied by the distribution used to fit the model. seen ⇓

2

In Figure 1, we see that the mean of the residuals is approximately 0 but the variance is greater than $\phi = 1$ as implied by part c) — therefore the variance of the response is larger than implied by the Poisson GLM. unseen ⇓ 2

seen ⇓

e) Plugging in an estimated dispersion parameter $\widehat{\phi} > 1$ will increase the standard errors.

1

sim. seen ⇓

More precisely, the standard errors will increase by a factor of $\sqrt{\widehat{\phi}_P}$. 2

Students have seen that $\mathrm{cov}(\widehat{\beta}) \approx \widehat{\phi}(X^T\widetilde{W}X)^{-1}$, *where* X *is the design matrix and* $\widetilde{W}$ *is the weight matrix from the IWLS algorithm. From this, they should be able to get the solution.*

4. a) The probability mass function for the Binomial distribution can be written as: seen $\Downarrow$

$$
\begin{aligned}
&\exp\left\{y\log(\pi)+(m-y)\log(1-\pi)+\log\binom{m}{y}\right\}\\
=&\exp\left\{y\log\left(\frac{\pi}{1-\pi}\right)+m\log(1-\pi)+\log\binom{m}{y}\right\},
\end{aligned}
$$

$\boxed{1}$

which is of the form given in Q3a). Therefore, we identify $\theta = \log(\pi/(1-\pi))$, $b(\theta) = -m\log(1-\pi) = m\log(1+\exp(\theta))$, $\phi = 1$, so $a(\phi) = 1$ and $c(y,\phi) = \log\binom{m}{y}$.

Students have seen exactly this just with a notion change; m for n. $\boxed{2}$

b) Following from a) the full log-likelihood, written in terms of $\boldsymbol{\mu} \equiv \mathrm{E}(\boldsymbol{Y})$, is

$$\ell(\boldsymbol{\mu};\boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i\log\left(\frac{\mu_i}{m_i-\mu_i}\right)+m_i\log(m_i-\mu_i)+\log\binom{m_i}{y_i}\right].$$

The maximum achievable log-likelihood, given by setting $\mu_i = y_i$ for all i: $\boxed{1}$

$$\ell(\boldsymbol{y};\boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i\log\left(\frac{y_i}{m_i-y_i}\right)+m_i\log(m_i-y_i)+\log\binom{m_i}{y_i}\right].$$

The maximised log-likelihood given by another model, with fitted values $\widehat{\mu}_i$, is $\boxed{1}$

$$\ell(\widehat{\boldsymbol{\mu}};\boldsymbol{y}) = \sum_{i=1}^{n}\left[y_i\log\left(\frac{\widehat{\mu}_i}{m_i-\widehat{\mu}_i}\right)+m_i\log(m_i-\widehat{\mu}_i)+\log\binom{m_i}{y_i}\right].$$

By the definition of the deviance, $\boxed{1}$

$$
\begin{aligned}
D =&2\phi\left\{\ell(\boldsymbol{y};\boldsymbol{y})-\ell(\widehat{\boldsymbol{\mu}};\boldsymbol{y})\right\}\\
=&2\sum_{i=1}^{n}\left\{y_i\log(y_i/\widehat{\mu}_i)+(m_i-y_i)\log\left(\frac{m_i-y_i}{m_i-\widehat{\mu}_i}\right)\right\}
\end{aligned}
$$

where we used $\phi = 1$ from a). $\boxed{2}$

c) Start by setting $m_i = 1$ in the deviance given in part b) $\boxed{\text{unseen} \Downarrow}$

$$D = 2\sum_{i=1}^{n}\left\{y_i \log(y_i/\widehat{\mu}_i) + (1-y_i)\log\left(\frac{1-y_i}{1-\widehat{\mu}_i}\right)\right\}$$

As $m_i = 1$, we have that $y_i \in \{0,1\}$. Now since both $y_i \log(y_i)$ and $(1-y_i)\log(1-y_i)$ are 0 for $y_i \in \{0,1\}$, the deviance reduces to $\boxed{1}$

$$\begin{aligned} D &= 2\sum_{i=1}^{n}\left\{y_i \log(1/\widehat{\mu}_i) + (1-y_i)\log\left(\frac{1}{1-\widehat{\mu}_i}\right)\right\} \\ &= -2\sum_{i=1}^{n}\left\{y_i \log\left(\frac{\widehat{\mu}_i}{1-\widehat{\mu}_i}\right) + \log(1-\widehat{\mu}_i)\right\} \end{aligned}$$

$\boxed{1}$

d) Differentiate to find

$$\begin{aligned} \frac{\partial}{\partial \beta_j}\ell(\beta; \boldsymbol{y}) &= \frac{\partial}{\partial \beta_j}\sum_{i=1}^{n}\left[y_i \log\left(\frac{\mu_i}{1-\mu_i}\right) + \log(1-\mu_i)\right] \\ &= \sum_{i=1}^{n}\left[\left(y_i\left\{\frac{1}{\mu_i} + \frac{1}{1-\mu_i}\right\} - \frac{1}{1-\mu_i}\right)\frac{\partial \mu_i}{\partial \beta_j}\right] \\ &= \sum_{i=1}^{n}\left[\left(\frac{y_i}{\mu_i(1-\mu_i)} - \frac{1}{1-\mu_i}\right)\frac{\partial \mu_i}{\partial \beta_j}\right] \end{aligned}$$

Now using the logit link function $\boxed{2}$

$$\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right) \implies \mu_i = \frac{1}{1+\exp(-\eta_i)} = \frac{1}{1+\exp(-\sum_j^p x_{ij}\beta_j)}$$

It follows that

$$\begin{aligned} \frac{\partial \mu_i}{\partial \beta_j} &= \frac{1}{(1+\exp(-\eta_i))^2}\exp(-\eta_i)x_{ij} \\ &= \mu_i(1-\mu_i)x_{ij} \end{aligned}$$

So $\boxed{2}$

$$\begin{aligned} \frac{\partial}{\partial \beta_j}\ell(\beta; \boldsymbol{y}) &= \sum_{i=1}^{n}\left[\left(\frac{y_i}{\mu_i(1-\mu_i)} - \frac{1}{1-\mu_i}\right)\mu_i(1-\mu_i)x_{ij}\right] \\ &= \sum_{i=1}^{n}\left[(y_i - \mu_i)x_{ij}\right] \end{aligned}$$

Next

$$\sum_{j=1}^{p} \beta_j \frac{\partial}{\partial \beta_j} \ell(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} \left[(y_i - \mu_i) \sum_{j=1}^{p} x_{ij} \beta_j \right]$$

$$= \sum_{i=1}^{n} \left[(y_i - \mu_i) \log \left(\frac{\mu_i}{1 - \mu_i} \right) \right] \quad \text{as} \quad \eta_i = \sum_{j=1}^{p} x_{ij} \beta_j$$

Finally, since the MLE $\widehat{\boldsymbol{\beta}}$ is such that $\frac{\partial}{\partial \beta_j} \ell(\widehat{\boldsymbol{\beta}}; \boldsymbol{y}) = 0$ it follows that $\boxed{2}$

$$0 = \sum_{i=1}^{n} \left[(y_i - \widehat{\mu}_i) \log \left(\frac{\widehat{\mu}_i}{1 - \widehat{\mu}_i} \right) \right]$$

$$\Longrightarrow \sum_{i=1}^{n} \left[y_i \log \left(\frac{\widehat{\mu}_i}{1 - \widehat{\mu}_i} \right) \right] = \sum_{i=1}^{n} \left[\widehat{\mu}_i \log \left(\frac{\widehat{\mu}_i}{1 - \widehat{\mu}_i} \right) \right]$$

Substituting this term into the deviance given in part c) gives the result. $\boxed{1}$

e) Notice that in part d) the deviance does not involve the observations y_i — therefore it is certainly not an appropriate measure of the goodness of fit between the observations and the fitted values. $\boxed{1}$ $\boxed{2}$