STAT 635 F2022 Assignment 2 - Due on Nov. 11, Friday, 2022, 11:59pm

Instructor: Xuewen Lu (This PDF is made from an R Markdown File)

2022-11-11

Contents

.

General Policies [Total 58 marks]

- Your assignment solutions should be created in `"RStudio"` with R Markdown to include the R codes and should be saved as an R Markdown file and a PDF file generated from the R Markdown file. You should name your files as `"Lastname-Firstname-Stat635-A2.Rmd"` and `"Lastname-Firstname-Stat635-A2.pdf"`.
- Test your R codes before submission to make sure it can be executed successfully in `"RStudio"`.
- For each assignment, submit only one PDF file and the associated Rmd file to D2L. Only the PDF file is graded. The Rmd file may be used to test your R programs when needed.
- If Monte-Carlo methods are used, you must fix the random seed in your R code by using **set.seed(2022)**.
- For all numerical problems, summarize the computer generated results in tables or figures and interpret them using your own words, then draw conclusions from them. **Show all your work and spell out the details**.
- Late submission is not acceptable.
- For guidelines on how to write a good assignment, go to D2L to read two sample assignments, **Bad-assignment-example.pdf** and **Good-assignment-example.pdf**, one is bad and the other is good, you are expected to do a good one.

Note: Materials are based on Dobson and Barnett (D&B or DB), 3rd Edition, 2008.

Problem 1. [6 marks]

Show for both Poisson regression and logistic regression that if an intercept is contained in the linear predictor then
$$\sum_{i=1}^{n}(y_i - \hat{\mu}_i) = 0.$$

Note: For both binomial and Poisson response y_i, they are counts instead of fractions.

Poisson Regression

Recall $Y \sim Poisson(\lambda_i)$ then $\mu_i = E[y_i] = \lambda_i$ and $\eta_i = g(\mu_i) = log(\mu_i) = x_i^T \beta = log(\lambda_i)$

Therefore, $exp(\eta_i) = \mu_i = \lambda_i$

Recall the likelihood function of a Poisson distribution

$$L(\lambda_i; y_i) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} = \prod_{i=1}^{n} exp(-\lambda_i + y_i log(\lambda_i) - log(y_i!))$$

We can write the corresponding log-likelihood function

$$l(\lambda_i; y_i) = log\Big(\prod_{i=1}^{n} exp(-\lambda_i + y_i log(\lambda_i) - log(y_i!)) \Big)$$
$$= \sum_{i=1}^{n} \Big(-\lambda_i + y_i log(\lambda_i) - log(y_i!) \Big)$$

Now rewrite it in terms of η_i

$$l(\eta_i; y_i) = \sum_{i=1}^{n} \Big(-exp(\eta_i) + y_i \eta_i - log(y_i!) \Big)$$

Let us differentiate the log-likelihood function with respect to η_i

$$\frac{\partial l(\eta_i; y_i)}{\partial \eta_i} = \sum_{i=1}^{n} \Big(y_i - exp(\eta_i) \Big)$$

Setting this to zero the MLE of η then satisfy

$$\sum_{i=1}^{n} \Big(y_i - exp(\hat{\eta}_i) \Big) = 0$$

Rearrange this equation and we can get

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} exp(\hat{\eta}_i) = \sum_{i=1}^{n} exp(x_i^T \hat{\beta}) = \sum_{i=1}^{n} \hat{\lambda}_i = \sum_{i=1}^{n} \hat{\mu}_i$$

This proves, in fact, $\sum_{i=1}^{n}(y_i - \hat{\mu}_i) = 0$

Logistic Regression

Recall when $Y_i \sim Bin(n_i, \pi_i)$, we have $\mu_i = E[y_i] = n_i \cdot \pi_i$ and $g(\mu_i) = \eta_i = x_i^T \beta = logit(\pi_i) = \frac{\pi_i}{1-\pi_i}$ and thus $expit(\eta_i) = \frac{exp(\eta_i)}{1+exp(\eta_i)} = \pi_i$

We denote $\frac{exp(\eta_i)}{1+exp(\eta_i)}$ as $\frac{b}{a}$ and keep in mind that $a - b = 1$ so that $1 - \frac{exp(\eta_i)}{1+exp(\eta_i)} = 1 - \frac{b}{a}$ and **Fact 1** states

$$(1 - \frac{b}{a})^{n-x} = (\frac{a-b}{a})^{n-x} = \frac{(a-b)^{n-x}}{a^{n-x}} = \frac{1}{a^{n-x}}$$

Recall the likelihood function of a binomial distribution

$$L(\pi_i; y_i) = \prod_{i=1}^{n} \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}$$

If we rewrite this likelihood function in terms of η_i then we have

$$L(\eta_i; y_i) = \prod_{i=1}^{n} \binom{n_i}{y_i} expit\left(\eta_i\right)^{y_i} \left(1 - expit(\eta_i)\right)^{n_i - y_i}$$

$$= \prod_{i=1}^{n} \binom{n_i}{y_i} \frac{exp(\eta_i)^{y_i}}{(1 + exp(\eta_i))^{y_i}} \frac{1^{n_i - y_i}}{(1 + exp(\eta_i))^{n_i - y_i}} \quad \text{by Fact 1 stated above}$$

$$= \prod_{i=1}^{n} \binom{n_i}{y_i} \frac{exp(\eta_i)^{y_i}}{(1 + exp(\eta_i))^{n_i}}$$

If we write the log likelihood function in terms of η_i and omit the constant term $\binom{n_i}{y_i}$ then we have

$$l(\eta_i; y_i) = \sum_{i=1}^{n} \left(y_i \cdot \eta_i - n_i \cdot log(1 + exp(\eta_i))\right)$$

We now differentiate the log likelihood function with respect to η_i

$$\frac{\partial l(\eta_i; y_i)}{\partial \eta_i} = \sum_{i=1}^{n} \left(y_i - n_i\left(\frac{exp(\eta_i)}{1 + exp(\eta_i)}\right)\right)$$

$$= \sum_{i=1}^{n} \left(y_i - n_i \cdot expit(\eta_i)\right)$$

Setting this to zero the maximum likelihood estimates of η then satisfy:

$$\sum_{i=1}^{n} \left(y_i - n_i \cdot expit(\hat{\eta}_i)\right) = 0$$

Rearrange the equation above and we can see

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} n_i \cdot expit(\hat{\eta}_i) = \sum_{i=1}^{n} n_i \cdot expit(x_i^T \hat{\beta}) = \sum_{i=1}^{n} n_i \cdot \hat{\pi}_i = \sum_{i=1}^{n} \hat{\mu}_i$$

This proves, in fact, $\sum_{i=1}^{n} (y_i - \hat{\mu}_i) = 0$

Problem 2. [8 marks, 2,2,2,2]

Let Y_i be the number of successes in n_i trials with

$$Y_i \sim Bin(n_i, \pi_i),$$

where the probabilities π_i have a Beta distribution

$$\pi_i \sim Be(\alpha, \beta).$$

The probability density function for the beta distribution is $f(x; \alpha, \beta) = x^{(\alpha-1)}(1-x)^{(\beta-1)}/B(\alpha, \beta)$ for x in $[0, 1]$, $\alpha > 0$, $\beta > 0$ and the beta function $B(\alpha, \beta)$ defining the normalizing constant required to ensure that $\int_0^1 f(x; \alpha, \beta)dx = 1$. Let $\theta = \alpha/(\alpha + \beta)$ and $\phi = 1/(\alpha + \beta + 1)$, show that

(a) $E(\pi_i) = \theta$.

Since $\pi_i \sim Be(\alpha, \beta)$., we compute the expected value of a Beta random variable X as:

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{+\infty} x f(x)dx \\
&= \int_0^1 x \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \int_0^1 \frac{1}{B(\alpha, \beta)} x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \qquad \text{by the integral representation of Beta Function} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha) \cdot a}{\Gamma(\alpha+\beta) \cdot (\alpha+\beta)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

So for a Beta-distributed random variable X, its expectation $E[X] = \frac{\alpha}{\alpha+\beta}$

Therefore, $E[\pi_i] = \frac{\alpha}{\alpha+\beta} = \theta$ as $\pi_i \sim Be(\alpha, \beta)$.

(b) $Var(\pi_i) = \phi\theta(1-\theta)$.

$E[X]^2 = \left(\frac{\alpha}{\alpha+\beta}\right)^2$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

$$= \int_0^1 x^2 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha+2-1}(1-x)^{\beta-1}dx$$

$$= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \qquad \text{by the integral representation of Beta Function}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)\cdot(\alpha+1)}{\Gamma(\alpha+\beta+1)\cdot(\alpha+\beta+1)}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha)\cdot a \cdot(a+1)}{\Gamma(\alpha+\beta)\cdot(\alpha+\beta)\cdot(\alpha+\beta+1)}$$

$$= \frac{\alpha\cdot(\alpha+1)}{(\alpha+\beta+1)\cdot(\alpha+\beta)}$$

$$Var[X] = E[X^2] - E[X]^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2}$$

$$= \frac{(\alpha+1)(\alpha+\beta)\alpha - (\alpha+\beta+1)\alpha^2}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

$$= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

$$= \phi\theta(1-\theta) \qquad \text{show as required}$$

(c) $E(Y_i) = n_i\theta$.

Note: $Y_i|\pi_i \sim Bin(n_i, \pi_i)$, so its expectation is $n_i\pi_i$ and variance is $n_i\pi_i(1-\pi_i)$

From part a), we know $E[\pi_i] = \frac{\alpha}{\alpha+\beta} = \theta$

$$E[Y_i] = E[E(Y_i|\pi_i)] = E[n_i\pi_i] = n_i \cdot \frac{\alpha}{\alpha+\beta} = n_i\theta$$

$$\text{proved by total expectation property}$$

(d) $Var(Y_i) = n_i\theta(1-\theta)[1+(n_i-1)\phi]$ so that $Var(Y_i)$ is larger than the Binomial variance (unless $n_i = 1$ or $\phi = 0$).

$Y_i|\pi_i \sim Bin(n_i, \pi_i)$, so its expectation is $n_i\pi_i$ and variance is $n_i\pi_i(1-\pi_i)$ and by the result of part a and b

$$Var[Y_i] = E[Var(Y_i|\pi_i)] + Var[E(Y_i|\pi_i)] = E[n_i\pi_i(1-\pi_i)] + Var[n_i\pi_i]$$

$$= n_i E[\pi_i] - n_i E[\pi_i^2] + n_i^2 Var[\pi_i]$$

$$= n_i\Big(E[\pi_i] - Var[\pi_i] - E[\pi_i]^2 + nVar[\pi_i]\Big)$$

$$= n_i \frac{\alpha\beta(\alpha+\beta+n_i)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$= n_i\theta(1-\theta)[1+(n_i-1)\phi] \qquad \text{proved by total variance property}$$

$$> n_i\theta(1-\theta)$$

6

Problem 3. [6 marks, 3,3]

In GLM with p covariates and an intercept, consider a situation in which a normal distribution is assumed with a log link.

(a) Describe the score function for maximum likelihood estimation of β, a $(p+1) \times 1$-vector of regression coefficients

With a log link function, in GLM with p covariates and an intercept, the general expression looks like this:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip})$$

Since y_i is normally distributed, we start with the likelihood function as always.

$$L(\mu, \sigma^2; y_i) = \prod_{i=1}^{n} f(y_i | \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right) \right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu)^2 \right)$$

Now we compute the log likelihood function

$$l(\mu, \sigma^2; y_i) = log\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu)^2 \right) \right)$$

$$= -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu)^2$$

Method 1: General Results for Exponential Family

Let's rewrite the likelihood function as a member of exponential family **(for a single observation y)**.

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

$$f(\theta, \phi; y) = exp\left(log\left(f(\theta, \phi; y) \right) \right) = exp\left(-\frac{1}{2} log(2\pi\sigma^2) - \frac{(y - \mu)^2}{2\sigma^2} \right)$$

$$= exp\left(\frac{-y^2 + 2y\mu - \mu^2}{2\sigma^2} - \frac{1}{2} log(2\pi\sigma^2) \right)$$

$$= exp\left(\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2} log(2\pi\sigma^2) \right)$$

Eventually, we rewrite it as a member of exponential family member where

$$\theta = \mu, b(\theta) = \frac{\mu^2}{2}, \phi = \sigma^2, \omega = 1, a(\phi) = \phi/\omega = \sigma^2$$

$$c(\phi; y) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}log(\frac{1}{2\pi\sigma^2}) = -\frac{1}{2}[log(2\pi\sigma^2) + y^2/\sigma^2]$$

Let's rewrite the log-likelihood function **(for a single observation y)**

$$l(\theta, \phi; y) = log(L(\theta, \phi; y)) = \frac{y * \theta - b(\theta)}{a(\phi)} + c(\phi; y)$$

So in this case, we have the log-likelihood function

$$\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}log(2\pi\sigma^2)$$

Recall: **for a single observation** y, we have $g(\mu) = \eta = x^T\beta = \beta_0 + \beta_1 x_1 + ...\beta_p x_p$

In log-linear model, $\mu = e^{x^T\beta} = e^\eta$

Now let us compute the jth entry of Score Vector $S(\beta)$, denoted as $[S(\beta)]_j$

$[S(\beta)]_j = \frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \theta}\frac{\partial \theta}{\partial \mu}\frac{\partial \mu}{\partial \eta}\frac{\partial \eta}{\partial \beta_j}$ where

$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} = \frac{y - \theta}{\sigma^2} = \frac{y - \mu}{\sigma^2}$, $\frac{\partial \theta}{\partial \mu} = \frac{1}{b''(\theta)} = 1$, $\frac{\partial \mu}{\partial \eta} = \frac{1}{g'(\mu)} = \mu = e^{x^T\beta} = e^\eta$ and $\frac{\partial \eta}{\partial \beta_j} = x_j$

Therefore, **for n observations$\{y_1, y_2, y_3, \ldots, y_n\}$ from the exponential family**, the score vector $S(\beta)$ can be described as

$$[S(\beta)]_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n}\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n}\frac{y_i - \theta_i}{\sigma_i^2}\cdot 1 \cdot e^{\eta_i}\cdot x_{ij} = \sum_{i=1}^{n}\frac{y_i - \mu_i}{\sigma_i^2}\cdot e^{x_i^T\beta}\cdot x_{ij}$$

In the form of Iterative Weighted Least Squares, $[S(\beta)]_j = \omega_i(z_i - \eta_i)x_{ij}$ where $\omega_i = \frac{1}{Var(Yi)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} = \frac{1}{\sigma_i^2(\frac{1}{\mu_i})^2}$

and $z_i = x_i^T\beta + (yi - \mu_i)(\frac{\partial \eta_i}{\partial \mu_i}) = \eta_i + (y_i - \mu_i)\cdot\frac{1}{\mu_i} = \eta_i + (y_i - \mu_i)\cdot\frac{1}{\theta_i} = x_i^T\beta + (y_i - e^{x_i^T\beta})\cdot\frac{1}{e^{x_i^T\beta}}$

Method 2: Direct Substitution Method

We have $g(\mu) = \eta = x^T\beta = \beta_0 + \beta_1 x_1 + ...\beta_p x_p$ and in log-linear model, $\mu = e^{x^T\beta} = e^\eta$

We implement a direct substitution to find the score function for maximum likelihood estimation of β

$$S(\beta) = \frac{\partial l(\mu)}{\partial \mu}\cdot\frac{\partial \mu}{\partial \eta}\cdot\frac{\partial \eta}{\partial \beta} = \frac{\partial l(\mu)}{\partial \mu}\cdot\frac{\partial \mu}{\partial e^\eta}\cdot\frac{\partial e^\eta}{\partial \beta}$$

For n observations $\{y_1, y_2, y_3, \ldots, y_n\}$

$\mu_i = E[Y_i]$ and Y_i is normally distributed, then it is quick to write down the likelihood function

$$L(\mu_i, \sigma_i^2; y_1, y_2, ...y_n) = (2\pi\sigma_i^2)^{\frac{-n}{2}}exp\left(-\frac{1}{2\sigma_i^2}\sum_{i=1}^{n}(y_i - \mu_i)^2\right)$$

The log-likelihood function can be written as

$$l(\mu_i, \sigma_i^2; y_1, y_2, ...y_n) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma_i^2) - \frac{1}{2\sigma_i^2}\sum_{i=1}^{n}(y_i - \mu_i)^2$$

8

Now let's compute $\frac{\partial l_i}{\partial \mu_i}$

$$\frac{\partial}{\partial \mu_i} l_i(\mu_i, \sigma_i^2; y_i) = \frac{\partial}{\partial \mu_i}\left(-\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma_i^2) - \frac{1}{2\sigma_i^2}\sum_{i=1}^{n}(y_i - \mu_i)^2\right) = \frac{1}{\sigma_i^2}\sum_{i=1}^{n}(y_i - \mu_i)$$

Please note $\frac{\partial l_i}{\partial \mu_i}$ can be re-written as $\frac{1}{\sigma_i^2}\sum_{i=1}^{n}(y_i - \mu_i) = \frac{1}{\sigma_i^2}\sum_{i=1}^{n}\left(y_i - exp(\eta_i)\right) = \frac{1}{\sigma_i^2}\sum_{i=1}^{n}\left(y_i - exp(x_i^T\beta)\right)$

Compute $\frac{\partial \mu_i}{\partial e^{\eta_i}} = \frac{\partial}{\partial e^{\eta_i}}e^{\eta_i} = 1$

Compute $\frac{\partial e^{\eta_i}}{\partial \beta} = \frac{\partial}{\partial \beta}e^{x_i^T\beta} = e^{x_i^T\beta}x_i$

So the Score Vector $[S(\beta)]_j = \sum_{i=1}^{n}\frac{\partial l_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial e^{\eta_i}}\frac{\partial e^{\eta_i}}{\partial \beta_j} = \sum_{i=1}^{n}\frac{y_i - \mu_i}{\sigma_i^2}e^{\eta_i}x_{ij}$

In conclusion, both Method 1 and Method 2 return the same result of Score Vector $S(\beta)$. One can solve $S(\beta)$=0 to get $\hat{\beta}$ or use other numeric methods to do the computation, such as Newton Raphson Method and Fisher Scoring Method or refer to Lecture Note 5

(b) Give the asymptotic covariance matrix for $\hat{\beta}_{MLE}$ and explain all terms in it.

By Lecture Note 5, asymptotic covariance matrix for $\hat{\beta}_{MLE}$ can be computed as the inverse of the Fisher Information Matrix such that $Cov(\hat{\beta}) = \Sigma_{\beta_{MLE}} = (X^TWX)^{-1}$ where $x_{ij} = \partial l_i/\partial \beta_j$ and $W = \frac{1}{Var(Y)}(\frac{\partial \mu}{\partial \eta})^2$

Since we have implemented log-linear model, so that $g(\mu_i) = log(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$

Therefore, $\mu_i = e^{\eta_i} = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}}$

W is the N $\times$ N diagonal matrix with elements $W_{ii} = \frac{1}{Var(Yi)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} = \frac{1}{\sigma_i^2(\frac{1}{\mu_i})^2} = \frac{\mu_i^2}{\sigma_i^2} = \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}}}{\sigma_i}\right)^2$

X is the N $\times$ (p+1) matrix as the following

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & ... & x_{2p} \\ ... & ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & x_{n3} & ... & x_{np} \end{pmatrix}$$

So the information matrix I can be expressed as:

$I = X^TWX$

$$= \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix}^T \begin{pmatrix} w_{11} & & \\ & \ddots & \\ & & w_{ii} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n}W_{ii} & \sum_{i=1}^{n}W_{ii}x_{i1} & ... & \sum_{i=1}^{n}W_{ii}x_{ip} \\ \sum_{i=1}^{n}W_{ii}x_{i1} & \sum_{i=1}^{n}W_{ii}x_{i1}^2 & ... & ... \\ ... & ... & ... & ... \\ \sum_{i=1}^{n}W_{ii}x_{ip} & \sum_{i=1}^{n}W_{ii}x_{i1}x_{ip} & ... & \sum_{i=1}^{n}W_{ii}x_{ip}^2 \end{pmatrix}$$

where $\sum_{i=1}^{n}W_{ii} = \sum_{i=1}^{n}\left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}}}{\sigma_i}\right)^2$, $\sum_{i=1}^{n}W_{ii}x_{ip} = \sum_{i=1}^{n}\left(\frac{e^{2\beta_0 + 2\sum_{j=1}^{p}\beta_j x_{ij}}}{\sigma_i^2}\right)x_{ip}$

where $\sum_{i=1}^{n}W_{ii}x_{ip}^2 = \sum_{i=1}^{n}\left(\frac{e^{2\beta_0 + 2\sum_{j=1}^{p}\beta_j x_{ij}}}{\sigma_i^2}\right)x_{ip}^2$ and so on, for i = 1,2,...n and for j = 1,2,...p

Note: I is a (p+1) by (p+1) matrix

Problem 4. [12 marks, 4,4,2,2]

The data in Table 1 show tumor responses of male and female patients receiving treatment for small-cell lung cancer. There were two treatment regimes. For the sequential treatment, the same combination of chemotherapeutic agents was administered at each treatment cycle. For the alternating treatment, different combinations were alternated from cycle to cycle (data from Holtbrugger and Schumacher 1991).

Table 1: Tumor responses to two different treatments: numbers of patients in each category.

Treatment	Sex	Progressive disease	No change	Partial remission	Complete remission
Sequential	Male	28	45	29	26
	Female	4	12	5	2
Alternating	Male	41	44	20	20
	Female	12	7	3	1

Note: when you define dummy variables, treat the first level as the reference level. So for `Treatment`, `Sequential` is the reference level; for `Sex`, `Male` is the reference level. For ordinal response, treat `Progressive disease` as the reference category. Use the R function `vglm()` in the R package `VGAM` and the ascending order in response category to do this problem.

(a) Fit a proportional odds model to estimate the probabilities for each response category taking treatment and sex effects into account.

For Ordinal Logistic Regression, we implement **Proportional Odds Model**

We assume J outcome categories are ordered and we are more interested in the cumulative response probabilities $\gamma_j(X_i) = P(Y \leq j)$ and we consider the odds of being in category j or lower:

$$log\Big(\frac{P(Y_i \leq j|X_i)}{1 - P(Y_i \leq j|X_i)}\Big) = log\Big(\frac{\gamma_j(X_i)}{1 - \gamma_j(X_i)}\Big) = X_i^T\beta_j = \beta_{0j} + \beta_1 X_{i1} + ... + \beta_p X_{ip}$$

Ascending order in response category and we have

$$L_1 = log(\frac{\pi_1}{\pi_2 + \pi_3 + \pi_4}), L_2 = log(\frac{\pi_1 + \pi_2}{\pi_3 + \pi_4}), L_3 = log(\frac{\pi_1 + \pi_2 + \pi_3}{\pi_4})$$

```
library(VGAM)
library(dplyr)
c_treatment = c(rep("sequential",2),rep("alternating",2))
c_sex = c("Male", "Female", "Male", "Female")
c_progressive = c(28,4,41,12)
c_nochange = c(45,12,44,7)
c_partial = c(29,5,20,3)
c_complete = c(26,2,20,1)


tumor.data = as.data.frame(cbind(c_treatment,c_sex,c_progressive,
                          c_nochange,c_partial,c_complete))
tumor.data = tumor.data %>% mutate_at(c('c_progressive', 'c_nochange','c_partial','c_complete'),
                          as.numeric)
head(tumor.data)
```

```
##   c_treatment   c_sex c_progressive c_nochange c_partial c_complete
## 1  sequential   Male             28         45        29         26
## 2  sequential Female              4         12         5          2
## 3 alternating   Male             41         44        20         20
## 4 alternating Female             12          7         3          1
```

```
library(reshape2)
tumor.data$c_treatment = factor(tumor.data$c_treatment,levels = c("sequential", "alternating"))
tumor.data$c_sex = factor(tumor.data$c_sex, levels = c("Male", "Female"))
```

```
cum.logit = vglm(cbind(c_progressive,c_nochange,c_partial,c_complete)~c_treatment+c_sex,
                 family = cumulative(parallel = TRUE),data = tumor.data)
summary(cum.logit)
```

```
##
## Call:
## vglm(formula = cbind(c_progressive, c_nochange, c_partial, c_complete) ~
##     c_treatment + c_sex, family = cumulative(parallel = TRUE),
##     data = tumor.data)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept):1            -1.3180     0.1801  -7.319  2.5e-13 ***
## (Intercept):2             0.2492     0.1621   1.538  0.12412
## (Intercept):3             1.3001     0.1852   7.021  2.2e-12 ***
## c_treatmentalternating    0.5807     0.2119   2.741  0.00613 **
## c_sexFemale               0.5414     0.2953   1.834  0.06671 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
## logitlink(P[Y<=3])
##
## Residual deviance: 5.5677 on 7 degrees of freedom
##
## Log-likelihood: -25.5417 on 7 degrees of freedom
##
## Number of Fisher scoring iterations: 5
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
## c_treatmentalternating            c_sexFemale
##               1.787262               1.718403
```

The fitted models look like:

$$L_1 = -1.318 + 0.581 \times Treatment + 0.541 \times Sex$$

$$L_2 = 0.249 + 0.581 \times Treatment + 0.541 \times Sex$$

$$L_3 = 1.300 + 0.581 \times Treatment + 0.541 \times Sex$$

(b) Examine the adequacy of the model fitted in (a) using residuals and goodness of fit statistics. [Hint: Use the McFadden's pseudo-R^2, the Pearson chi-squared goodness of fit statistic X^2 and the deviance D of the fitted model as shown in the Example for the car preferences to do this problem.]

$$\text{Pseudo-}R^2 = \frac{Null.Dev - Resid.Dev}{Null.Dev} = \frac{Dev(b_{min}) - Dev(b)}{Dev(b_{min})} = \frac{l(b_{min}) - l(b))}{l(b_{min}) - l(b_{max}))}$$

```
dev_b = deviance(cum.logit) #Dev(b)
cum.logit_null = vglm(cbind(c_progressive,c_nochange,c_partial,c_complete)~1,
               family = cumulative(parallel = TRUE),data = tumor.data)
dev_n = deviance(cum.logit_null) #Dev(bmin)
Pseudo_R_squared = (dev_n-dev_b)/dev_n
Pseudo_R_squared
```

```
## [1] 0.6621351
```

$$\text{Pearson's } \chi^2 = X^2 = \sum (\text{Pearson Residuals})^2$$

```
sum(resid(cum.logit,"pearson")^2)
```

```
## [1] 5.352753
```

$$\text{Deviance } D = 2\Big(l(b_{max}) - l(b)\Big)$$

```
dev = deviance(cum.logit)
dev
```

```
## [1] 5.567678
```

```
criteria = c("McFadden Pseudo R^2", "Pearson Chi Squared", "Deviance")
value = c(0.6621351,5.352753,5.567678)
table = data.frame(criteria,value)
knitr::kable(table, "pipe", col.names = c("Criteria", "Value"), align = c("c","l"))
```

Criteria	Value
McFadden Pseudo R^2	0.6621351
Pearson Chi Squared	5.3527530
Deviance	5.5676780

(c) Use a Wald statistic to test the hypothesis that there is no difference in responses for the two treatment regimes.

Step 1: $H_0 : \beta_{treatment} = 0$ and $H_a : \beta_{treatment} \neq 0$

Step 2: Wald-test Statistics $\frac{(\hat{\beta}_j - 0)^2}{\hat{Var}(\hat{\beta}_j)} = \frac{0.5807^2}{0.2119^2} = 7.513081 \sim \chi_1^2$ same as $z = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{0.5807}{0.2119} = 2.741 \sim N(0,1)$

Step 3: $p - value = 2 \times P(Z > |z|) = 0.00613 < 0.05$ where $Z \sim N(0,1)$

Step 4: We reject H_0 at a significance level of 5%. This means that there is a difference in responses for the two treatment regimes.

(d) Fit two proportional odds models to test the hypothesis of no treatment difference. Compare the results with those for (c) above.

```
cum.logit2 = vglm(cbind(c_progressive,c_nochange,c_partial,c_complete)~+c_sex,
                  family = cumulative(parallel = TRUE),data = tumor.data)
summary(cum.logit2)
```

```
##
## Call:
## vglm(formula = cbind(c_progressive, c_nochange, c_partial, c_complete) ~
##      +c_sex, family = cumulative(parallel = TRUE), data = tumor.data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -1.0150     0.1383  -7.337 2.18e-13 ***
## (Intercept):2   0.5188     0.1277   4.063 4.84e-05 ***
## (Intercept):3   1.5571     0.1605   9.704  < 2e-16 ***
## c_sexFemale     0.5219     0.2940   1.775   0.0759 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
## logitlink(P[Y<=3])
##
## Residual deviance: 13.1379 on 8 degrees of freedom
##
## Log-likelihood: -29.3268 on 8 degrees of freedom
##
## Number of Fisher scoring iterations: 4
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
## c_sexFemale
##    1.685159
```

So in this case, the model we proposed in part a is considered as a saturated model and the model we proposed in part d is considered as a restricted model. So we perform likelihood ratio test for differences in models.

```
1-pchisq(deviance(cum.logit2) - deviance(cum.logit), df.residual(cum.logit2)-df.residual(cum.logit))
```

```
## [1] 0.00593417
```

Since this p-value is less than 0.05, so we reject the null hypothesis, meaning that the restricted model proposed in part d is not adequate as compared to the saturated model proposed in part a.

Problem 5. [14 marks, 4,2,2,2,2,2]

A study was performed to investigate new automobile purchases. A sample of 20 families was selected. Each family was surveyed to determine the age of their oldest vehicle and their total family income. A follow-up survey was conducted 6 months later to determine if they had actually purchased a new vehicle during that time period ($y = 1$ indicates yes and $y = 0$ indicates no). The data from this study are shown in the file `vehicle.csv`. You can read the data using the following R chunk,

```
options(digits=6) #show more decimal values;
library(ggplot2) #for use of ggplot()
library(knitr)   #for use of kable()

##Set up the data, first change the format of Income in the dataset from characters to numerical values
# vehicle=read.table("C://Users/elham/OneDrive/Desktop/A 2/vehicle.txt",header = TRUE)
vehicle=read.table("vehicle.txt",header = TRUE)
vehicle = data.frame("Income" = c(as.numeric(vehicle$Income)),"Age" = c(vehicle$Age),"y" = c(vehicle$y)
head(vehicle)
```

```
##   Income Age y
## 1  45000   2 0
## 2  40000   4 0
## 3  60000   3 1
## 4  50000   2 1
## 5  55000   2 0
## 6  50000   5 1
```

(a) Fit a logistic regression model to the data and state your model.

Logistic Model: $logit\big(p(y_i = 1)\big) = log\left(\frac{p(y_i=1)}{1-p(y_i=1)}\right) = \eta_i = x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ for $i = 1, 2, ...20$ where $\hat{\beta}_0 = -7.05$, $\hat{\beta}_1 = 7.38e - 05$ and $\hat{\beta}_2 = 9.88e - 01$

```
model = glm(y~Income+Age, family = binomial(link = "logit"),data = vehicle)
summary(model)
```

```
##
## Call:
## glm(formula = y ~ Income + Age, family = binomial(link = "logit"),
##     data = vehicle)
##
## Deviance Residuals:
##    Min     1Q  Median     3Q    Max
## -1.563  -0.804  -0.140  0.954  1.792
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.05e+00   4.67e+00   -1.51    0.132
## Income       7.38e-05   6.37e-05    1.16    0.247
## Age          9.88e-01   5.27e-01    1.87    0.061 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

14

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 27.726  on 19  degrees of freedom
## Residual deviance: 21.082  on 17  degrees of freedom
## AIC: 27.08
##
## Number of Fisher Scoring iterations: 5
```

(b) Does the model deviance indicate that the logistic regression model from part (a) is adequate?

H_0 : The proposed model with 2 parameters is adequate

H_a : The proposed model with 2 parameters is not adequate as compared to the saturated model

p-value $= P[\chi^2_{17} > D]$

If p-value $< \alpha$, then there is evidence against H_0. Else, we do not reject H_0

```
1-pchisq(model$deviance,17)
```

```
## [1] 0.222665
```

Therefore, we do not reject H_0, which means that the proposed model from part a is adequate

(c) Interpret the model coefficients β_1 and β_2.

$\hat{\beta}_1 = 7.38e - 05$: holding variable *age* at a fixed value, we will see 0.007% increase in the odds of purchasing a new car for a one-unit increase in *income* since $\exp(7.38e-05) = 1.00007$

$\hat{\beta}_2 = 9.88e - 01$: holding variable *income* at a fixed value, we will see 168.586% increase in the odds of purchasing a new car for a one-unit increase in *age* since $\exp(9.88e-01) = 2.68586$

(d) What is the estimated probability that a family with an income of \$45,000 and a car that is 5 years old will purchase a new vehicle in the next 6 months?

$p(y = 1) = expit(-7.05 + 7.38 \times 10^{-5} \times 45000 + 0.988 \times 5) = 0.770476$

```
(exp(-7.05+7.38*10^(-5)*45000+0.988*5))/(1+exp(-7.05+7.38*10^(-5)*45000+0.988*5))
```

```
## [1] 0.770476
```

(e) Expand the linear predictor to include an interaction term. Is there any evidence that this term is required in the model?

$logit\big(p(y_i = 1)\big) = log\left(\frac{p(y_i=1)}{1-p(y_i=1)}\right) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2}$

```
model2 = glm(y~Income*Age, family = binomial(link = "logit"),data = vehicle)
summary(model2)
```

```
##
## Call:
## glm(formula = y ~ Income * Age, family = binomial(link = "logit"),
##     data = vehicle)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -1.6398   -0.6275   -0.0564    0.6621    1.8567
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.314351   6.393966    0.05     0.96
## Income      -0.000141   0.000141   -1.00     0.32
## Age         -2.461695   2.081492   -1.18     0.24
## Income:Age   0.000101   0.000063    1.61     0.11
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 27.726  on 19   degrees of freedom
## Residual deviance: 16.551  on 16   degrees of freedom
## AIC: 24.55
##
## Number of Fisher Scoring iterations: 6
```

```
## delta D
model$deviance - model2$deviance
```

```
## [1] 4.53072
```

```
# p-value
1-pchisq(model$deviance - model2$deviance,1)
```

```
## [1] 0.0332917
```

This means that the reduced/nested model is adequate. The model with interaction term is not required in the model.

(f) Find approximate 95% confidence intervals on the model parameters for the logistic regression model from part (a).

CI for β_0 is $\hat{\beta}_0 \pm 1.96 se(\hat{\beta}_0) = -7.05 \pm 1.96 \times 4.67 : (-16.2, 2.1)$

CI for β_1 is $\hat{\beta}_1 \pm 1.96 se(\hat{\beta}_1) = 0.0000738 \pm 1.96 \times 0.0000637 : (-5.1052e-05, 1.98652e-04)$

CI for β_2 is $\hat{\beta}_2 \pm 1.96 se(\hat{\beta}_2) = 0.988 \pm 1.96 \times 0.527 : (-0.04492, 2.02092)$

Problem 6 [12 marks, 2,2,2,2,2,2]

The Aircraft Damage Data: During the Vietnam War, the United States Navy operated several types of attack (a bomber in USN parlance) aircraft, often for low-altitude strike missions against bridges, roads, and other transportation facilities. Two of these included the McDonnell Douglas A-4 Skyhawk and the Grumman A-6 Intruder. The A-4 is a single-engine, single place light-attack aircraft used mainly in daylight. It was also flown by the Blue Angels, the Navy's flight demonstration team, for many years.The A-6 is a twin-engine, dual-place, all-weather medium-attack aircraft with excellent day/night capabilities. However, the Intruder could not be operated from the smaller Essex class aircraft carriers, many of which were still in service during the conflict. Considerable resources were deployed against the A-4 and A-6, including small arms, AAA or antiaircraft artillery, and surface-to-air missiles. The following table contains data from 30 strike missions involving these two types of aircraft. The regressor x_1 is an indicator variable (A-4=0 and A-6=1), and the other regressors x_2 and x_3 are bomb load (in tons) and total months of aircrew experience. The response variable y is the number of locations where damage was inflicted on the aircraft.

We model the damage response as a function of the three regressors. Since the response is a count, we use a Poisson regression model with the log link. Read data using the following R chunk,

```
options(digits=6, width=200)
air<-read.csv("aircraft-data.csv")
head(air)
```

```
##   Observation y x1 x2   x3
## 1           1 0  0  4 91.5
## 2           2 1  0  4 84.0
## 3           3 0  0  4 76.5
## 4           4 0  0  5 69.0
## 5           5 0  0  5 61.5
## 6           6 0  0  5 80.0
```

Do the following:

(a) Fit a full model with three predictors. State your fitted model.

```
model1 = glm(y~x1+x2+x3, family=poisson(link=log), data=air)
model1$coefficients
```

```
## (Intercept)          x1          x2          x3
##  -0.4060227   0.5687724   0.1654254  -0.0135223
```

$$log(\hat{\lambda}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3}$$

$$log(\hat{\lambda}_i) = -0.40602 + 0.56877 x_{i1} + 0.16543 x_{i2} - 0.01352 x_{i3}$$

(b) To investigate the potential usefulness of various subset models, fit all three two-variable models and all three one-variable models to the data and provide a summary of the results obtained in a table: create a 4-columns table, with titles `Model`, `Deviance`, `Difference in Deviance Compared to Full Model`, `p-value`, respectively.

```r
model2_1 = glm(y~x1+x2, family=poisson(link=log), data=air)
model2_2 = glm(y~x1+x3, family=poisson(link=log), data=air)
model2_3 = glm(y~x2+x3, family=poisson(link=log), data=air)
model3_1 = glm(y~x1, family=poisson(link=log), data=air)
model3_2 = glm(y~x2, family=poisson(link=log), data=air)
model3_3 = glm(y~x3, family=poisson(link=log), data=air)


#deviance
model2_1.d = glm(y~x1+x2, family=poisson(link=log), data=air)$deviance
model2_2.d = glm(y~x1+x3, family=poisson(link=log), data=air)$deviance
model2_3.d = glm(y~x2+x3, family=poisson(link=log), data=air)$deviance
model3_1.d = glm(y~x1, family=poisson(link=log), data=air)$deviance
model3_2.d = glm(y~x2, family=poisson(link=log), data=air)$deviance
model3_3.d = glm(y~x3, family=poisson(link=log), data=air)$deviance
#df
model2_1.df = glm(y~x1+x2, family=poisson(link=log), data=air)$df.residual
model2_2.df = glm(y~x1+x3, family=poisson(link=log), data=air)$df.residual
model2_3.df = glm(y~x2+x3, family=poisson(link=log), data=air)$df.residual
model3_1.df = glm(y~x1, family=poisson(link=log), data=air)$df.residual
model3_2.df = glm(y~x2, family=poisson(link=log), data=air)$df.residual
model3_3.df = glm(y~x3, family=poisson(link=log), data=air)$df.residual


Model = c("x1,x2","x1,x3","x2,x3", "x1", "x2", "x3")
Deviance = c(model2_1.d,model2_2.d,model2_3.d,
             model3_1.d,model3_2.d,model3_3.d)
Res.Deviance = c(model2_1.d-model1$deviance,model2_2.d-model1$deviance,model2_3.d-model1$deviance,
                 model3_1.d-model1$deviance,model3_2.d-model1$deviance,model3_3.d-model1$deviance)
df.Deviance = c(model2_1.df-model1$df.residual,model2_2.df-model1$df.residual,
                model2_3.df-model1$df.residual,model3_1.df-model1$df.residual,
                model3_2.df-model1$df.residual,model3_3.df-model1$df.residual)
p.value = c(1-pchisq(Res.Deviance[1], df.Deviance[1]),1-pchisq(Res.Deviance[2], df.Deviance[2]),
            1-pchisq(Res.Deviance[3], df.Deviance[3]),1-pchisq(Res.Deviance[4], df.Deviance[4]),
            1-pchisq(Res.Deviance[5], df.Deviance[5]),1-pchisq(Res.Deviance[6], df.Deviance[6]))


table = data.frame(Model,Deviance,Res.Deviance,p.value)
knitr::kable(table, "pipe",
             col.names = c("Model",
                           "Deviance",
                           "Difference in Deviance Compared to Full Model",
                           "p-value"),
             align = c("l","l","c","l"))
```

Model	Deviance	Difference in Deviance Compared to Full Model	p-value
x1,x2	28.6343	2.68115	0.101543
x1,x3	32.1917	6.23859	0.012500
x2,x3	27.2198	1.26667	0.260393
x1	38.2835	12.33030	0.002101
x2	29.2059	3.25273	0.196643
x3	50.5372	24.58408	0.000005

(c) Base on the deviance tests in (b), at a test size $\alpha = 10\%$, are these subset models different from the full model? What is the best model you recommend among these models?

18

They are all different from the full model, but two of them (two-variable model that contains x1 and x2 and another two-variable model that contains x2 and x3) are adequate as compared to the full model. I would suggest the two-variable model that contains x2 and x3 and consider this model to be the best after the analysis of Deviance Table

(d) Based on the full model, use the Wald test to find the most non-significant covariate. After removing the most non-significant covariate, fit a sub-model using the rest two covariates, is there any non-significant covariate in these two covariates? If yes, remove it and fit a one covariate model and test its significance using the Wald test.

```
model1 = glm(y~x1+x2+x3, family=poisson(link=log), data=air)
summary(model1)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3, family = poisson(link = log),
##     data = air)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.642   -1.006   -0.018    0.558    1.909
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.40602    0.87749   -0.46    0.644
## x1           0.56877    0.50437    1.13    0.259
## x2           0.16543    0.06754    2.45    0.014 *
## x3          -0.01352    0.00828   -1.63    0.102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 53.883  on 29  degrees of freedom
## Residual deviance: 25.953  on 26  degrees of freedom
## AIC: 87.65
##
## Number of Fisher Scoring iterations: 5
```

This shows that x_1 is the most non-significant covaraite.

```
#only x2 and x3 are in the proposed model this time
modelp = glm(y~x2+x3, family=poisson(link=log), data=air)
summary(modelp)
```

```
##
## Call:
## glm(formula = y ~ x2 + x3, family = poisson(link = log), data = air)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5670   -0.9860   0.0453    0.3279   2.1989
```

19

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.69883    0.85302   -0.82     0.41
## x2           0.22228    0.04646    4.78  1.7e-06 ***
## x3          -0.01174    0.00845   -1.39     0.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 53.883  on 29  degrees of freedom
## Residual deviance: 27.220  on 27  degrees of freedom
## AIC: 86.92
##
## Number of Fisher Scoring iterations: 5
```

This shows that x_3 is the most non-significant covaraite.

```
#only x2 is contained in the proposed model this time
modelpp = glm(y~x2, family=poisson(link=log), data=air)
summary(modelpp)
```

```
##
## Call:
## glm(formula = y ~ x2, family = poisson(link = log), data = air)
##
## Deviance Residuals:
##    Min     1Q  Median     3Q     Max
## -1.919  -1.047  -0.152  0.165   2.633
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.7010     0.5069   -3.36  0.00079 ***
## x2            0.2311     0.0468    4.94  7.7e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 53.883  on 29  degrees of freedom
## Residual deviance: 29.206  on 28  degrees of freedom
## AIC: 86.9
##
## Number of Fisher Scoring iterations: 5
```

Wald Test

Step 1: $H_0 : \beta = 0$ and $H_a : \beta \neq 0$

Step 2: Wald Statistic:

$$W = \frac{(\hat{\beta} - 0)^2}{I(\hat{\beta})^{-1}} = \frac{0.2311^2}{0.0468^2} = 24.3842 \sim \chi^2_{(df=1)}$$

This is equivalent to

$$z = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{0.2311}{0.0468} = 4.93803 \sim N(0,1)$$

Step 3: from the summary table above, we see $p - value = 2 \times P(Z > |z|) = 0.00000077 < 0.05$ where $Z \sim N(0,1)$

Step 4: we reject the null hypothesis, which means that this variable x_2 (boom load in tons) is statistically significant.

(e) Now you obtain two models from (c) and (d), are these two models the same or different? If there are different, what caused the difference? Finally, which model do you want to keep?

No, they are different. I would say that different techniques of variable selection may result in different answers. In part C, we use Deviance Test, in part D, we use backward selection. According to the table in part b), I would keep the two variable model, which contains x_2 and x_3 since it has a smallest deviance difference.

(f) Using the sign of estimated regression coefficients in your final model to interpret their effects on the response? Does the interpretation make practical sense?

```
model2_3 = glm(y~x2+x3, family=poisson(link=log), data=air)
summary(model2_3)
```

```
##
## Call:
## glm(formula = y ~ x2 + x3, family = poisson(link = log), data = air)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.5670  -0.9860   0.0453   0.3279   2.1989
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.69883    0.85302   -0.82     0.41
## x2           0.22228    0.04646    4.78  1.7e-06 ***
## x3          -0.01174    0.00845   -1.39     0.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 53.883  on 29  degrees of freedom
## Residual deviance: 27.220  on 27  degrees of freedom
## AIC: 86.92
##
## Number of Fisher Scoring iterations: 5
```

x_2 represents bomb load (in tons) and x_3 represents total months of aircrew experience. The response variable y is the number of locations where damage was inflicted on the aircraft. The sign of $\hat{\beta}_2$ is negative, which makes sense. The more bomb loads the aircraft can carry, the more damage it can make. The sign of $\hat{\beta}_3$ is negative, which does not make sense. Ideally, the more experienced the aircrew is, the more damage it can make. But it is close zero, so this does not make a huge change afterall.