1. (a)

\[ X^T R = X^T(Y - \hat{Y}) \]
\[ = X^T(Y - X\hat{\beta}) \]
\[ = X^T(Y - X(X^TX)^{-1}X^TY) \]
\[ = X^TY - X^TX(X^TX)^{-1}X^TY \]
\[ = X^TY - X^TY = 0. \]

(b) The product \( X^T R \) is a vector with \( j \)-th entry equal to \( \sum_i x_{ij} R_i \), where \( x_{ij} \) denotes the \((i, j)\) entry of \( X \). If (say) the first column of \( X \) is a column of 1’s, that is \( x_{i1} = 1 \) for all \( i \), then the first entry of \( X^T R \) is \( \sum_i x_{i1} R_i = \sum_i R_i \). Since \( X^T R = 0 \), it follows that \( \sum_i R_i = 0 \).

2. If \( Z_1, Z_2 \) are i.i.d. Normal \( N(0, 1) \) then the vector \( Z = (Z_1, Z_2) \) has a Multivariate Normal Distribution with mean vector \( \mu = (0, 0) \) and variance-covariance matrix

\[ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \]

Now if

\[ Y_1 = Z_1 + Z_2 \]
\[ Y_2 = Z_1 - Z_2 \]

then the vector \( Y = (Y_1, Y_2) \) satisfies

\[ Y = AZ \]

where

\[ A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \]

It follows (by a theorem stated in lectures) that \( Y \) has a Multivariate Normal Distribution with mean vector \( A\mu = A(0, 0) = (0, 0) \) and variance-covariance matrix

\[ A\Sigma A^T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \]
\[ = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \]

Thus, \( Y_1 \) and \( Y_2 \) are independent Normal random variables with mean 0 and variance 2.

3. (a)

\[ E[\hat{\beta}] = E[(X^TV^{-1}X)^{-1}X^TV^{-1}Y] \]
\[ = (X^TV^{-1}X)^{-1}X^TV^{-1}E[Y] \]
\[ = (X^TV^{-1}X)^{-1}X^TV^{-1}X\beta \]
\[ = \beta. \]
(b) The variance-covariance matrix of a random vector \( \mathbf{Z} \) is \( \text{var}(\mathbf{Z}) = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])^T] \). Here we want the variance-covariance matrix of \( \hat{\beta} = AY \) where \( A = (X^TV^{-1}X)^{-1}X^TV^{-1} \).

First,

\[
\hat{\beta} - \mathbb{E}[\hat{\beta}] = AY - \mathbb{E}[AY] = A(Y - \mathbb{E}[Y])
\]

so that

\[
\text{var}(\hat{\beta}) = \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = \mathbb{E}[(A(Y - \mathbb{E}[Y])(A(Y - \mathbb{E}[Y]))^T] = A\mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T]A^T.
\]

The last expectation is the variance-covariance matrix of \( Y \), which equals \( aV \). Thus

\[
\text{var}(\hat{\beta}) = AaVA^T = aAV^T = a(X^TV^{-1}X)^{-1}X^TV^{-1}YV^{-1}X(X^TV^{-1}X)^{-1} = a(X^TV^{-1}X)^{-1}.
\]

(c) Since \( \hat{\beta} = AY \) and \( Y \) is multivariate normal, the random vector \( \hat{\beta} \) has a Multivariate Normal distribution with mean vector \( \mathbb{E}[\hat{\beta}] = \beta \) and variance-covariance matrix \( a(X^TV^{-1}X)^{-1} \).

4. (a) This is a linear model in which the design matrix \( X \) is an \( n \times 1 \) matrix with entries \( x_1, \ldots, x_n \), the parameter vector \( \beta \) has a single entry \( \alpha \), and the errors have variance-covariance matrix

\[
c = \begin{pmatrix}
x_1 & 0 & \ldots & 0 \\
0 & x_2 & \ldots & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \ldots & x_n
\end{pmatrix}.
\]

Letting \( V \) be the diagonal matrix with entries \( x_1, \ldots, x_n \), this model is covered by the results of the previous question. Thus the maximum likelihood estimator \( \hat{\alpha} = \hat{\beta} \) is given by

\[
\hat{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}Y = (\sum_i x_i)^{-1}(\sum_i y_i) = \frac{\sum_i y_i}{\sum_i x_i}.
\]

(b) By the results of the previous question, the distribution of \( \hat{\alpha} \) is Normal with mean \( \alpha \) and variance \( \text{var}(\hat{\alpha}) = c(X^TV^{-1}X)^{-1} = c/\sum_i x_i \).

5. (a) The likelihood is

\[
L(\beta_1, \ldots, \beta_q) = \prod_{i=1}^n f_i(y_i, \beta_1, \ldots, \beta_q) = \prod_{i=1}^n (1 - p_i)^{y_i-1} p_i = \prod_{i=1}^n (1 - \frac{1}{\mu_i})^{y_i-1} \frac{1}{\mu_i}.
\]
(b) The likelihood for the fitted model is, from part (a),

$$L(\hat{\mu}_1, \ldots, \hat{\mu}_n) = \prod_{i=1}^{n} (1 - \frac{1}{\hat{\mu}_i})^{y_i-1} \frac{1}{\hat{\mu}_i}$$

while for the saturated model the likelihood is

$$L(\hat{\mu}_1, \ldots, \hat{\mu}_n) = \prod_{i=1}^{n} (1 - \frac{1}{\mu_i})^{y_i-1} \frac{1}{\mu_i}$$

$$= \prod_{i=1}^{n} (1 - \frac{1}{y_i})^{y_i-1} \frac{1}{y_i}.$$  

Hence the deviance is

$$\Lambda = 2 \log \frac{L(\hat{\mu}_1, \ldots, \hat{\mu}_n)}{L(\hat{\mu}_1, \ldots, \hat{\mu}_n)}$$

$$= 2 \sum_{i=1}^{n} [(y_i - 1) \log \frac{1 - 1/y_i}{1 - 1/\hat{\mu}_i} - \log \hat{\mu}_i + \log y_i].$$

6. (a)

$$L(\beta_1, \ldots, \beta_n) = \prod_{i=1}^{n} f_i(y_i, \beta_1, \ldots, \beta_n)$$

$$= \prod_{i=1}^{n} e^{-\beta_i y_i}.$$  

(b) For the fitted model the likelihood is

$$L(\hat{\mu}_1, \ldots, \hat{\mu}_n) = \prod_{i=1}^{n} e^{-\hat{\mu}_i \frac{y_i}{y_i!}}$$

while for the saturated model ($\hat{\mu}_i = y_i$) the likelihood is

$$L(\hat{\mu}_1, \ldots, \hat{\mu}_n) = \prod_{i=1}^{n} e^{-\hat{\mu}_i \frac{y_i}{y_i!}}$$

$$= \prod_{i=1}^{n} e^{-y_i \frac{y_i}{y_i!}}.$$  

Hence the deviance is

$$D = 2 \log \frac{L(\hat{\mu}_1, \ldots, \hat{\mu}_n)}{L(\hat{\mu}_1, \ldots, \hat{\mu}_n)}$$

$$= \sum_{i=1}^{n} [(-y_i + y_i \log y_i - \log(y_i!)) - (-\hat{\mu}_i + y_i \log \hat{\mu}_i - \log(y_i!))]$$

$$= \sum_{i=1}^{n} [(\hat{\mu}_i - y_i) + y_i \log \frac{y_i}{\hat{\mu}_i}].$$