1. (40 points) Consider the Galapagos islands data with different geological variables measured from 30 islands. For example, the row for the island Baltra is as follows:

```
#               Species    Area Elevation Nearest Scruz Adjacent
# Baltra             58   25.09       346     0.6   0.6     1.84
```

Consider the relationship between the number of species (counts) and these geological variables. The Poisson regression with the log link yields the following output.

```
Call:
glm(formula = Species ~ ., family = poisson, data = gala)

Deviance Residuals:
Min       1Q   Median       3Q      Max
-8.2752  -4.4966  -0.9443   1.9168  10.1849

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.155e+00  5.175e-02  60.963  < 2e-16 ***
Area        -5.799e-04  2.627e-05 -22.074  < 2e-16 ***
Elevation    3.541e-03  8.741e-05  40.507  < 2e-16 ***
Nearest      8.826e-03  1.821e-03   4.846 1.26e-06 ***
Scruz       -5.709e-03  6.256e-04  -9.126  < 2e-16 ***
Adjacent    -6.630e-04  2.933e-05 -22.608  < 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: 3510.73  on 29  degrees of freedom
Residual deviance:  716.85  on 24  degrees of freedom
AIC: 889.68

Number of Fisher Scoring iterations: 5
```

Based on the output, answer the following questions.

(a) What is the interpretation of the slope for "Area"?

(b) What is the fitted value $\hat{\mu}_{Baltra}$ for the island Baltra?

(c) What are the Pearson residual and deviance residual for the island Baltra?

(d) What is the sum of all response residuals? Why?

2. (28 points) A random vector $\boldsymbol{Y}_{J\times 1} = (Y_1, \ldots, Y_J)$ is said to follow a multinomial distribution, denoted by $Multinomial(n, \boldsymbol{\pi})$, where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$ with $\sum_{j=1}^J \pi_j = 1$, if

$$f(\boldsymbol{y}) = \frac{n!}{\prod_{j=1}^J y_j!} \prod_{j=1}^J \pi_j^{y_j}.$$

Show the following properties of $Multinomial(n, \boldsymbol{\pi})$.

(a) $Multinomial(n, \boldsymbol{\pi})$ is a $(J-1)$-dimensional exponential family.

(b) $\forall j \neq k$, $Cov(Y_j, Y_k) = -n\pi_j\pi_k$.

(c) $\forall j$, $(Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_J)|Y_j = y_j \sim Multinomial(n - y_j, (\frac{\pi_1}{1-\pi_j}, \ldots, \frac{\pi_J}{1-\pi_j}))$.

(d) Suppose $W_1, \ldots, W_j$ are independent with $W_j \sim Poisson(\lambda_j)$, where $\lambda_j > 0$ and $j = 1, \ldots, J$. Then

$$(W_1, \ldots, W_J)|\sum_{j=1}^J W_j \sim Multinomial(\sum_{j=1}^J W_j, (\frac{\lambda_1}{\sum_{j=1}^J \lambda_j}, \ldots, \frac{\lambda_J}{\sum_{j=1}^J \lambda_j})).$$

3. (32 points) Consider the linear mixed model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ where $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times p}$ with $rank(\mathbf{X}) = p < n$, $\mathbf{Z} \in \mathbb{R}^{n\times q}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\alpha} \sim \mathcal{N}_q(\mathbf{0}, \sigma_\alpha^2\mathbf{I})$ is a random vector of i.i.d. coefficients, and $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\epsilon^2\mathbf{I})$ is a vector of i.i.d. random errors and is independent of $\boldsymbol{\alpha}$.

Consider the REML with $\mathbf{K}_{(n-p)\times n}$ matrix s.t. $rank(\mathbf{K}) = n - p$ and $\mathbf{K}\mathbf{X} = \mathbf{0}_{(n-p)\times p}$. We mentioned in class that the choice of $\mathbf{K}$ is not unique, and different choices of $\mathbf{K}$ lead to the same estimate. Let's work out why here.

(a) Show that for any positive definite matrix $\mathbf{V}$,

$$\mathbf{K}^T(\mathbf{K}\mathbf{V}\mathbf{K}^T)^{-1}\mathbf{K} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}.$$

(Hint: Note that $\mathbf{K}\mathbf{X} = \mathbf{K}\mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{X} = \mathbf{0}_{(n-p)\times p}$.)

(b) Write the likelihood $\mathcal{L}(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2|\mathbf{K}\boldsymbol{y})$ in terms of $\mathbf{X}$, $\mathbf{Z}$, and $\boldsymbol{y}$ (but without $\mathbf{K}$!). (Hint: You may assume (a) here. Also use Sylvester's determinant theorem: For matrices $\mathbf{A}_{s\times t}$ and $\mathbf{B}_{t\times s}$, $det(\mathbf{I}_s + \mathbf{A}\mathbf{B}) = det(\mathbf{I}_t + \mathbf{B}\mathbf{A})$)