1. Consider the following data set ratstime.txt:

```
  time poison treat
1 0.31      I     A
2 0.82      I     B
3 0.43      I     C
4 0.45      I     D
5 0.45      I     A
6 1.10      I     B
.
.

Effect of toxic agents on rats

Description

An experiment was conducted as part of an investigation
to combat the effects of certain toxic agents.

A data frame with 48 observations on the following 3 variables.

time
survival time in tens of hours

poison
the poison type - a factor with levels I II III

treat
the treatment - a factor with levels A B C D
```

The response variable is $Y =$ time and the explanatory variables are X_1=poison and X_2=treat.

(a) Model the data with the main effect model

$$\mu_{jh} = \beta_0 + \beta_j + \alpha_h,$$

with the assumption of $Y_{ijk} \sim N(\mu_{jh}, \sigma^2)$. Test all pairwise average differences

$$H_0 : \left(\frac{\mu_{1h} + \mu_{2h} + \cdots + \mu_{kh}}{k}\right) - \left(\frac{\mu_{1h_*} + \mu_{2h_*} + \cdots + \mu_{kh_*}}{k}\right) = 0, \qquad h \neq h_*,$$

i.e., test whether there is average differences on means in the levels of X_2=`treat` variable. Particularly, report the Wald test statistic value obtained from the comparison of A and B treatments.

(2 points)

(b) Model the data with the main effect model

$$\mu_{jh} = \beta_0 + \beta_j + \alpha_h,$$

with the assumption of $Y_{ijk} \sim N(\mu_{jh}, \sigma^2)$. Test all pairwise predictive average differences

$$H_0 : \left(\frac{Y_{i1h} + Y_{i2h} + \cdots + Y_{ikh}}{k}\right) - \left(\frac{Y_{i1h_*} + Y_{i2h_*} + \cdots + Y_{ikh_*}}{k}\right) = 0, \qquad h \neq h_*,$$

i.e., test whether there is average differences on random variables in the levels of X_2=`treat` variable. Particularly, report the so-called d-value obtained from the comparison of A and B treatments.

(2 points)

(c) Let us continue to assume $Y_i \sim N(\mu_{jh}, \sigma^2)$. Model the data with the models

$$\begin{aligned} \mu_{jh} &= \beta_0 + \beta_j + \alpha_h, \\ \log(\mu_{jh}) &= \beta_0 + \beta_j + \alpha_h, \\ \frac{1}{\mu_{jh}} &= \beta_0 + \beta_j + \alpha_h. \end{aligned}$$

Calculate the mean square error value $\mathrm{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_{jh})^2}{n}$ for all these models. Which link function fits best to the data?

(2 points)

2. Consider the data in file Alba.txt.

```
> data<-read.table("Alba.txt", sep="\t", dec=".", header=TRUE)
> head(data)
  Dose  Herbicide DryMatter
1    1 Glyphosate       4.7
2    1 Glyphosate       4.6
3    1 Glyphosate       4.1
4    1 Glyphosate       4.4

Data are from an experiment, comparing the potency of the two herbicides
glyphosate and bentazone in white mustard Sinapis alba.
Dose - a numeric vector containing the dose in g/ha.
Herbicide - a factor with levels Bentazone Glyphosate (the two herbicides applied).
DryMatter - a numeric vector containing the response (dry matter in g/pot).
Christensen, M. G. and Teicher, H. B., and Streibig, J. C. (2003)
Linking fluorescence induction curve and biomass in herbicide screening,
Pest Management Science, 59, 1303?1310.
```

Denote the variables as $Y =$ DryMatter, $X_1 =$ Dose, and $X_2 =$ Herbicide.

(a) Let us assume $Y_i \sim Gamma(\mu_i, \phi)$. Consider modeling the expected value μ_i of the response variable $Y =$ DryMatter by the following model

$$\mathcal{M}: \quad \frac{1}{\mu_i} = \beta_0 + \beta_1 x_{i1} + \alpha_j + \gamma_j x_{i1},$$

where index j is related to the categories of X_2. Calculate the maximum likelihood estimate for the expected value μ_{i_*} when $X_1 = 50$ and $X_2 =$ Glyphosate. (2 points)

(b) Let us continue to assume $Y_i \sim Gamma(\mu_i, \phi)$, and let us continue to model the expected value μ_i by the model

$$\mathcal{M}: \quad \frac{1}{\mu_i} = \beta_0 + \beta_1 x_{i1} + \alpha_j + \gamma_j x_{i1},$$

Create 80 % prediction interval for new observation y_f, when $X_1 = 50$ and $X_2 =$ Glyphosate. Particularly, what is your estimate for lowerbound of the prediction interval? (2 points)

(c) Let us assume $Y_i \sim IG(\mu_i, \phi)$. Consider modeling the expected value μ_i of the response variable $Y =$ DryMatter by the following model

$$\mathcal{M}: \quad \frac{1}{\mu_i^2} = \beta_0 + \beta_1 x_{i1} + \alpha_j + \gamma_j x_{i1},$$

where index j is related to the categories of X_2. Calculate the 95% confidence interval estimate for the expected value μ_{i_*} when $X_1 = 50$ and $X_2 =$ Glyphosate. Particularly, what is your estimate for lowerbound of the confidence interval? (2 points)

3. (a) Let us assume $Y_i \sim IG(\mu_i, \phi)$. Consider the model

$$\log(\mu_i) = \beta_0 + \beta_1 \log(x_i).$$

Let the estimates of the parameters β_0, β_1, ϕ be as $\hat{\beta}_0 = 1, \hat{\beta}_1 = 0.5, \tilde{\phi} = 0.05$.

i. Calculate the maximum likelihood estimate for the expected value μ_i when $x_i = 5$.

ii. Calculate the Pearson residual

$$o_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\mathrm{Var}}(Y_i)}},$$

when $x_{i_*} = 5$ and the observed value is $y_i = 12$.

(2 points)

(b) Let $Y_i \sim Poi(\mu_i)$. Then the probability density function of the random variable Y_i is

$$f(y_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}.$$

Show first that Y_i belongs to the exponential family of distributions, and then show that

$$\mathrm{E}(Y_i) = \mu_i,$$
$$\mathrm{Var}(Y_i) = \mu_i.$$

Hint! There is no dispersion parameter ϕ in Poisson distribution and hence you may consider it as $\phi = 1$.

(2 points)

(c) In generalized linear models, the likelihood equations can written in form

$$\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mathrm{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = 0, \qquad j = 0, 1, 2 \ldots p.$$

Consider now the simple Gamma model with

$$Y_i \sim Gamma(\mu_i, \phi),$$
$$\mu_i = \eta_i = \beta_0.$$

What kind of more simplified form the likelihood equations have in this case? That is, what form $\frac{\partial l(\beta_0)}{\partial \beta_0}$ has in the simple Gamma model? By using the likelihood equations, find the maximum likelihood estimator $\hat{\beta}_0$.

(2 points)