A1 The inverse Gaussian distribution has density of the form

$$f(y;\nu,\lambda) = \left(\frac{\lambda}{2\pi y^3}\right)^{1/2} \exp\left(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right)$$

for $y > 0$, with parameters $\mu > 0$ and $\lambda > 0$.

(a) Show that the family of inverse Gaussian distributions is an exponential dispersion family. Identify the functions $b(\cdot)$, $c(\cdot)$ as well as the canonical and the dispersion parameters.

(b) Compute the mean and the variance of an inverse Gaussian random variable Y and identify the mean-variance relationship.

(c) Identify the canonical link for a GLM with inverse Gaussian responses. Do you think this link is sensible? What other link functions might be appropriate?

(d) Consider a GLM with inverse Gaussian responses and the canonical link. Write down the likelihood equations.

(e) Write down the likelihood equations in the special case where $g(\mu_i) = \beta_0 + \beta_1 x_i$ with $x_i = 1$ for $i = 1, \ldots, n_A$ from group A and $x_i = 0$ for $i = n_A + 1, \ldots, n_A + n_B$ from group B (here, $n = n_A + n_B$). Calculate the fitted means for groups A and B.

A2 Generalize your finding from **A1** (e): Show that for any link function and any GLM of the form $g(\mu_i) = \beta_0 + \beta_1 x_i$ with $x_i = 1$ for $i = 1, \ldots, n_A$ from group A and $x_i = 0$ for $i = n_A + 1, \ldots, n_A + n_B$ from group B (here, $n = n_A + n_B$), the fitted means $\hat{\mu}_A$ and $\hat{\mu}_B$ equal to the sample group means $\bar{y}_A$ and $\bar{y}_B$, respectively, where

$$\bar{y}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} y_i, \quad \bar{y}_B = \frac{1}{n_B}\sum_{n_A+1}^{n_A+n_B} y_i.$$

A3 (a) Show that an alternative expression for the GLM score equations is

$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{\operatorname{var}(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \quad j = 1, \ldots, p.$$

Show that these equations result from the generalized least squares problem of minimizing

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\operatorname{var}(y_i)},$$

treating the variances as known constants.

(b) For a GLM with canonical link function and $a(\phi)$ independent of i, explain how the score equations imply that the residual vector $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{\mu}}$ is orthogonal to the column space of X.

A4 `R` **excercise.** Load `Beetles2.dat` available on MyCourses under Assigmnents:

```
beetles <- read.table("Beetles2.dat",header=TRUE)
attach(beetles)
```

This data, from Bliss (1935, *Ann. Appl. Biol.*), shows the number of `dead` beetles out of `n` after 5 hours of exposure to gaseous carbon disulphide at different dosages (dosage is reported on the log scale, viz. `logdosage`).

(a) Fit an appropriate GLM using the function `glm` with the canonical link. Print the `summary` of the fit and the estimated parameters. Is `logdosage` a significant preditor?

(b) Interpret the effect of `logdosage` in the model from part (a). (Hint: consider odds and odds ratios).

(c) Think of two other link functions that would be appropriate and fit the corresponding GLMs again using `glm`.

(d) Construct a plot to assess how well the models from parts (a) and (b) fit.

(e) For the GLMs in parts (a) and (b), compute the fitted number of dead beetles at each considered level of `logdosage`.

(f) Out of the GLMs from parts (a) and (b), select the model that is most suitable for the data at hand using a suitable criterion. Justify your choice.