Question 3

If the random variable $Y$ has the Inverse Gaussian distribution with a parameter of interest $\mu$ and $\lambda$, then its probability density function is given by

$$f(y) = \left[ \frac{\lambda}{2\pi y^3} \right]^{1/2} \exp \left( \frac{-\lambda(y - \mu)^2}{2\mu^2 y} \right).$$  

(a) Show that the distribution of $Y$ belongs to the exponential family.

(b) What is the canonical link?

(c) Using the general results of exponential family show that the expectations of $Y$, $E(Y) = \mu$ and the variance of $Y$, $Var(y) = \mu^3/\lambda$.

(d) Based on the mean variance relationship above, when would you use an inverse gaussian distribution to model your glm.

An alternative way of modeling such a mean variance relationship is to design a quasi-likelihood starting from the mean variance relations $V(\mu) = \mu^3$.

(e) Using quasi-score $U = U(\mu; Y) = \frac{Y - \mu}{\sigma^2 V(\mu)}$ for the above variance function construct the quasi-log-likelihood.

(f) Show that with $\sigma^2 = 1/\lambda$ the above quasi likelihood is proportional to the inverse gaussian distribution given in (1).
**Question 4**

This question is based on the analysis of the following study:

The data record details the Birth to Ten study (BTT), performed in the greater Johannesburg-/Soweto metropolitan area of South Africa during 1990. In the study, all mothers of singleton births were interviewed during a seven-week period between April and June to women with permanent addresses in a defined area (a total of 4019 births). Five years later, 964 of these mothers were re-interviewed. If the mothers interviewed later are representative of the original populations, the two groups should show similar characteristics. Of those characteristics we will analyze the proportion with and without medical aid.

There are eight observations on four variables.

- **Counts:** The number of subjects in the given classification
- **Group:** Which group the mother belonged to:
  - 1 refers to the mothers not followed up after the five years;
  - 2 refers to the mothers followed-up five years later.
- **MedicalAid:** Whether or not the mother had medical aid; either Yes or No
- **Race:** The mothers’ race; either White or Black

The Full dataset is given below

<table>
<thead>
<tr>
<th>Counts</th>
<th>Group</th>
<th>MedicalAid</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>104</td>
<td>1</td>
<td>White</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>2</td>
<td>White</td>
</tr>
<tr>
<td>3</td>
<td>91</td>
<td>1</td>
<td>Black</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>2</td>
<td>Black</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>1</td>
<td>White</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>White</td>
</tr>
<tr>
<td>7</td>
<td>957</td>
<td>1</td>
<td>Black</td>
</tr>
<tr>
<td>8</td>
<td>368</td>
<td>2</td>
<td>Black</td>
</tr>
</tbody>
</table>

Alternative representation of the data in contingency table format

```r
> xtabs(~MedicalAid+Race+Group, data=btt)

Race
MedicalAid Black White
No     957 22
Yes    91 104

, , Group = 2

Race
MedicalAid Black White
No     368 2
Yes    36 10
```
(a) Analyze the interaction plot to comment on the pairwise interaction of the covariates Group, MedicalAid and Race.

We will naturally use a Poisson model to analyze this contingency table.

(b) Should we use offset to analyze our data? If so what will be our offset.

To analyze the data two possible models are included in the output. Model 1 is a additive model and Model 2 includes interaction between Race and MedicalAid. Page 1 uses use the regular poisson glm, where as page 2 accommodates for overdispersion by using the quasipoisson family.

(c) Based on the R output choose between the simple poisson glm models and the quasipoisson glm models and justify your choice. Use your chosen set of models for analyzing the rest of the question.

(d) First we analyze the dataset to answer the claim “If the mothers interviewed later are representative of the original populations, the two groups should show similar characteristics.”
(e) Based on the additive model which covariates are significant?

(f) Based on the interaction model which covariates are significant?

(g) Explain your findings based on the significant covariates of the two models.

(h) Compare the two models based on their deviance values and the appropriate anova table.

(i) Without using statistical terms provide a summary of your findings based on your chosen model.

---

**OUTPUT STARTS HERE**

---

**R output for Model 1 (additive model) and Model 2 (including interaction between Race and MedicalAid) fusing the regular poisson glm**

```
Call:
  glm(formula = Counts ~ Group + MedicalAid + Race, family = poisson, data = btt)

Coefficients:  
                Estimate Std. Error z value Pr(>|z|)  
(Intercept)    7.85050   0.07735   101.49  <2e-16 ***  
Group           1.03749   0.05706    18.18  <2e-16 ***  
MedicalAidYes  -1.72232   0.06993   -24.63  <2e-16 ***  
RaceWhite      -2.35344   0.08908   -26.42  <2e-16 ***  

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2886.81 on 7 degrees of freedom  
Residual deviance: 393.05 on 4 degrees of freedom
```

```
Call:
  glm(formula = Counts ~ Group + MedicalAid * Race, family = poisson, data = btt)

Coefficients:  
                Estimate Std. Error z value Pr(>|z|)  
(Intercept)    7.92334   0.07705   102.83  <2e-16 ***  
Group           1.03749   0.05706    18.18  <2e-16 ***  
MedicalAidYes  -2.34498   0.09289   -25.24  <2e-16 ***  
RaceWhite      -4.01111   0.20596   -19.48  <2e-16 ***  
MedicalAidYes:RaceWhite  3.90313   0.24304    16.06  <2e-16 ***  

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2886.810 on 7 degrees of freedom  
Residual deviance: 29.142 on 3 degrees of freedom
```

```
> anova(btt.glm,btt.glm.int,test="Chi")
Analysis of Deviance Table

Model 1: Counts ~ Group + MedicalAid + Race
Model 2: Counts ~ Group + MedicalAid * Race
              Resid. Df Resid. Dev  Deviance P(>|Chi|)  
1                  4        393.05          1   
2                  3         29.14  363.91 3.968e-81
```

---
R output for Model 1 (additive model) and Model 2 (including interaction between Race and MedicalAid) using the poisson with overdispersion [family quasipoisson]

Call:
  glm(formula = Counts ~ Group + MedicalAid + Race, family = quasipoisson, data = btt)
Coefficients:

  Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.8505     0.9745    8.056  0.00129 **
  Group       -1.0375     0.7188   -1.443  0.22241
MedicalAidYes -1.7223     0.8810   -1.955  0.12226
RaceWhite     -2.3534     1.1222   -2.097  0.10399

(Dispersion parameter for quasipoisson family taken to be 158.7084)

Null deviance: 2886.81 on 7 degrees of freedom
Residual deviance: 393.05 on 4 degrees of freedom

Call:
  glm(formula = Counts ~ Group + MedicalAid * Race, family = quasipoisson, data = btt)
Coefficients:

  Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.9233     0.2174   36.440  4.55e-05 ***
  Group       -1.0375     0.1610   -6.443  0.00758 **
MedicalAidYes -2.3450     0.2621   -8.946  0.00295 **
RaceWhite     -4.0111     0.5812  -6.901  0.00623 **
MedicalAidYes:RaceWhite  3.9031     0.6858   5.691  0.01076 *

(Dispersion parameter for quasipoisson family taken to be 7.963434)

Null deviance: 2886.810 on 7 degrees of freedom
Residual deviance: 29.142 on 3 degrees of freedom

> anova(btt$glm.quasi,btt$glm.int.quasi,test="Chi")
Analysis of Deviance Table
Model 1: Counts ~ Group + MedicalAid + Race
Model 2: Counts ~ Group + MedicalAid * Race
     Resid. Df Resid. Dev Df Deviance P(>Chi)
1           4     393.05
2           3      29.14  1  363.91 1.38e-11

> anova(btt$glm.quasi,btt$glm.int.quasi,test="F")
Analysis of Deviance Table
Model 1: Counts ~ Group + MedicalAid + Race
Model 2: Counts ~ Group + MedicalAid * Race
     Resid. Df Resid. Dev Df Deviance  F  Pr(>F)
1           4     393.05
2           3      29.14  1     363.91 45.697 0.006614 **