Qualifying Exam: CAS MA 576.

Boston University, Spring 2009

Question 3

Let $\{Y_1, ..., Y_n\}$ Be a set of independent RVs with $E(Y_i) = \mu_i$, $var(Yi) = \sigma^2 V(\mu_i)$, And

$$V(\mu) = \mu(1 + \alpha\mu)$$

for some known constant α.

(a) Show that the quasi-likelihood (QL) for a single Y is

$$y \log \frac{\mu}{y} - \left(y + \frac{1}{\alpha}\right) \log \frac{(1 + \alpha\mu)}{(1 + \alpha y)}$$

> Hint: Write down the Quasi Score U. Show that
>
> $$\int \frac{y - t}{t(1 + \alpha t)} = y \log(t) - \left(y + \frac{1}{\alpha}\right) \log(1 + \alpha t),$$
>
> If you want you can use the identity
>
> $$\frac{1}{t(1 + \alpha t)} = \frac{1 + \alpha t - \alpha t}{t(1 + \alpha t)} = \frac{1}{t} - \frac{\alpha}{1 + \alpha t}$$

(b) Calculate the QL for $\{Y_1, ..., Y_n\}$ and derive the quasi-deviance (QD).

(c) Write down the QD for $\alpha = -1$. What exponential family does the QD correspond to in this case?

(d) Show that when $\alpha = 0$, the QD is the same as the deviance for Poisson distributed data.

(e) Can you obtain the deviance under the normal distribution and gamma distribution as special cases of the above $V(\mu)$? If yes for which values of α would they correspond to. Otherwise propose a variance function $V(mu)$, which may include Normal, Gamma, Binomial and Poisson as special cases. [You do not need to work out the QL/QD]

Question 4

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

The data are taken from a clinical trial of 59 epileptics, who were randomized to receive either the anti-epileptic drug progabide or a placebo. The data on each subject consist of a baseline count of the number of seizures over an eight-week period prior to treatment allocation, followed by counts in four two-week periods post-treatment.

```
var     Description
----------------------------------------------------------------------
y1      counts of number of seizures in first two-week period
        after post-treatment

y2      counts of number of seizures in second two-week period
        after post-treatment

y3      counts of number of seizures in third two-week period
        after post-treatment

y4      counts of number of seizures in fourth two-week period
        after post-treatment

trt     whether received anti-epileptic drug progabide (trt=1) or placebo (trt=0)

base    baseline count of the number of seizures over an eight-week
        period prior to treatment allocation

age     age of the subject
----------------------------------------------------------------------
```

In this question we shall analyze a subset of data taken from a clinical trial of 59 epileptics, who were randomized to receive either the anti-epileptic drug progabide or a placebo. The data on each subject consist of a baseline count of the number of seizures over an eight-week period prior to treatment allocation, followed by counts in four two-week periods post-treatment.

The following few pages provides outputs from a range of Poisson models with and without overdispersion for modeling the number of seizures in the fourth two-week period of post-treatment (variable y4), in terms of the baseline count, age and treatment. Describe the steps for selecting the best model commenting specifically and justifying your choice for using

(a) Overdispersion Poisson vs Standard poisson regression.

(b) Inclusion and exclusion of interaction terms (2nd and 3rd order).

(c) Inclusion and exclusion of main effects.

(d) Model comparison based on "F" vs "Chisq" distribution.

Write down the statistical model that you finally choose and justify your choice.

Description of Models

```
seiz.mult=glm(y4~ trt*base*age,data=seizure,family=poisson)
seiz.twoway=glm(y4~ trt*base+trt*age+age*base,data=seizure,family=poisson)
seiz.add=glm(y4~ trt+base+age,data=seizure,family=poisson)
seiz.base.trt=glm(y4~ base*trt,data=seizure,family=poisson)
seiz.onlybase=glm(y4~ base,data=seizure,family=poisson)

seiz.mult.overd=glm(y4~ trt*base*age,data=seizure,family=quasipoisson)
seiz.twoway.overd=glm(y4~ trt*base+trt*age+age*base,data=seizure,family=quasipoisson)
seiz.add.overd=glm(y4~ trt+base+age,data=seizure,family=quasipoisson)
seiz.base.trt.overd=glm(y4~ base*trt,data=seizure,family=quasipoisson)
seiz.onlybase.overd=glm(y4~ base,data=seizure,family=quasipoisson)
```

Summary of Possion Glm Models

```
##################################################################################################
> summary(seiz.mult)
Call:
glm(formula = y4 ~ trt * base * age, family = poisson, data = seizure)

Coefficients:
                Estimate Std. Error z value P(>|z|)
(Intercept)    1.3400322  0.7148376   1.875   0.0608 .
trt           -0.0243711  0.8832661  -0.028   0.9780
base           0.0072530  0.0157319   0.461   0.6448
age           -0.0027286  0.0240156  -0.114   0.9095
trt:base      -0.0160298  0.0194789  -0.823   0.4105
trt:age       -0.0209968  0.0309935  -0.677   0.4981
base:age       0.0004560  0.0005262   0.867   0.3862
trt:base:age   0.0009598  0.0007311   1.313   0.1893
---
    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 134.77  on 51  degrees of freedom
AIC: 339.03

##################################################################################################
> summary(seiz.twoway)

Call:
glm(formula = y4 ~ trt * base + trt * age + age * base, family = poisson,
    data = seizure)

Coefficients:
               Estimate Std. Error z value P(>|z|)
(Intercept)  1.9306546  0.5469775   3.530 0.000416 ***
trt         -0.9594936  0.5062628  -1.895 0.058060 .
base        -0.0077995  0.0108607  -0.718 0.472670
age         -0.0228912  0.0184028  -1.244 0.213538
trt:base     0.0091700  0.0035132   2.610 0.009050 **
trt:age      0.0137299  0.0154728   0.887 0.374888
base:age     0.0009635  0.0003608   2.670 0.007581 **
---

    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 136.51  on 52  degrees of freedom
AIC: 338.76
```

```
################################################################################
> summary(seiz.add)

Call:
glm(formula = y4 ~ trt + base + age, family = poisson, data = seizure)

Coefficients:
             Estimate Std. Error z value P(>|z|)
(Intercept)  0.728587   0.245920   2.963  0.00305 **
trt         -0.278916   0.098866  -2.821  0.00479 **
base         0.022153   0.001092  20.293  < 2e-16 ***
age          0.015612   0.007159   2.181  0.02921 *
---


    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 144.57  on 55  degrees of freedom
AIC: 340.83

################################################################################
> summary(seiz.base.trt)

Call:
glm(formula = y4 ~ base * trt, family = poisson, data = seizure)


Coefficients:
              Estimate Std. Error z value P(>|z|)
(Intercept)  1.2530157  0.1172758  10.684   <2e-16 ***
base         0.0208717  0.0019112  10.920   <2e-16 ***
trt         -0.3606720  0.1619755  -2.227   0.0260 *
base:trt     0.0008258  0.0022749   0.363   0.7166
---


    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 149.04  on 55  degrees of freedom
AIC: 345.30

################################################################################
> summary(seiz.onlybase)

Call:
glm(formula = y4 ~ base, family = poisson, data = seizure)

Coefficients:
             Estimate Std. Error z value P(>|z|)
(Intercept) 1.0987359  0.0770681   14.26   <2e-16 ***
base        0.0208148  0.0009973   20.87   <2e-16 ***
---


    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 159.40  on 57  degrees of freedom
AIC: 351.66
```

Summary of Possion Glm Models using overdispersion

```
> summary(seiz.twoway.overd)
Call:
glm(formula = y4 ~ trt * base + trt * age + age * base, family = quasipoisson,
                         data = seizure)

Coefficients:
               Estimate Std. Error t value P(>|t|)
(Intercept)  1.9306546  0.8579587   2.250   0.0287 *
trt         -0.9594936  0.7940958  -1.208   0.2324
base        -0.0077995  0.0170355  -0.458   0.6490
age         -0.0228912  0.0288657  -0.793   0.4314
trt:base     0.0091700  0.0055106   1.664   0.1021
trt:age      0.0137299  0.0242698   0.566   0.5740
base:age     0.0009635  0.0005660   1.702   0.0947 .
---

(Dispersion parameter for quasipoisson family taken to be 2.460333)

    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 136.51  on 52  degrees of freedom
##################################################################################
> summary(seiz.add.overd)
Call:
glm(formula = y4 ~ trt + base + age, family = quasipoisson, data = seizure)

Coefficients:
             Estimate Std. Error t value P(>|t|)
(Intercept)  0.728587   0.383258   1.901   0.0625 .
trt         -0.278916   0.154080  -1.810   0.0757 .
base         0.022153   0.001701  13.021   <2e-16 ***
age          0.015612   0.011158   1.399   0.1673
---

(Dispersion parameter for quasipoisson family taken to be 2.428819)

    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 144.57  on 55  degrees of freedom
##################################################################################
> summary(seiz.base.trt.overd)
Call:
glm(formula = y4 ~ base * trt, family = quasipoisson, data = seizure)

Coefficients:
               Estimate Std. Error t value P(>|t|)
(Intercept)  1.2530157  0.1876564   6.677 1.27e-08 ***
base         0.0208717  0.0030582   6.825 7.30e-09 ***
trt         -0.3606720  0.2591817  -1.392    0.170
base:trt     0.0008258  0.0036401   0.227    0.821
---

(Dispersion parameter for quasipoisson family taken to be 2.560414)

    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 149.04  on 55  degrees of freedom
##################################################################################
```

```
> summary(seiz.onlybase.overd)

Call:
glm(formula = y4 ~ base, family = quasipoisson, data = seizure)

Coefficients:
             Estimate Std. Error t value P(>|t|)
(Intercept) 1.098736   0.124966   8.792 3.38e-12 ***
base        0.020815   0.001617  12.872  < 2e-16 ***
---

(Dispersion parameter for quasipoisson family taken to be 2.629268)

    Null deviance: 473.79  on 58  degrees of freedom
Residual deviance: 159.40  on 57  degrees of freedom
```

Comaprison of Possion Glm Models

```
################################################################################
> anova(seiz.mult,seiz.twoway,seiz.add,seiz.onlybase,test="Chisq")
Analysis of Deviance Table

Model 1: y4 ~ trt * base * age
Model 2: y4 ~ trt * base + trt * age + age * base
Model 3: y4 ~ trt + base + age
Model 4: y4 ~ base
  Resid. Df Resid. Dev Df Deviance  P(>|Chi|)
1        51    134.770
2        52    136.507 -1   -1.737     0.187
3        55    144.569 -3   -8.062     0.045
4        57    159.403 -2  -14.834     0.001

################################################################################
> anova(seiz.twoway,seiz.base.trt,seiz.onlybase,test="Chisq")
Analysis of Deviance Table

Model 1: y4 ~ trt * base + trt * age + age * base
Model 2: y4 ~ base * trt
Model 3: y4 ~ base
  Resid. Df Resid. Dev Df Deviance   P(>|Chi|)
1        52    136.507
2        55    149.038 -3  -12.531     0.006
3        57    159.403 -2  -10.365     0.006
```

Comparison of Possion Glm Models using overdispersion

```
> anova(seiz.twoway.overd,seiz.add.overd,seiz.onlybase.overd,test="Chisq")
Analysis of Deviance Table

Model 1: y4 ~ trt * base + trt * age + age * base
Model 2: y4 ~ trt + base + age
Model 3: y4 ~ base
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        52    136.507
2        55    144.569 -3   -8.062     0.351
3        57    159.403 -2  -14.834     0.049
#################################################################################
> anova(seiz.twoway.overd,seiz.base.trt.overd,seiz.onlybase.overd,test="Chisq")
Analysis of Deviance Table

Model 1: y4 ~ trt * base + trt * age + age * base
Model 2: y4 ~ base * trt
Model 3: y4 ~ base
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        52    136.507
2        55    149.038 -3  -12.531     0.165
3        57    159.403 -2  -10.365     0.122
#################################################################################
> anova(seiz.twoway.overd,seiz.add.overd,seiz.onlybase.overd,test="F")
Analysis of Deviance Table

Model 1: y4 ~ trt * base + trt * age + age * base
Model 2: y4 ~ trt + base + age
Model 3: y4 ~ base
  Resid. Df Resid. Dev Df Deviance      F  >F)
1        52    136.507
2        55    144.569 -3   -8.062 1.0922 0.36069
3        57    159.403 -2  -14.834 3.0146 0.05772 .


#################################################################################
> anova(seiz.twoway.overd,seiz.base.trt.overd,seiz.onlybase.overd,test="F")
Analysis of Deviance Table

Model 1: y4 ~ trt * base + trt * age + age * base
Model 2: y4 ~ base * trt
Model 3: y4 ~ base
  Resid. Df Resid. Dev Df Deviance      F >F)
1        52    136.507
2        55    149.038 -3  -12.531 1.6977 0.1789
3        57    159.403 -2  -10.365 2.1063 0.1319
#################################################################################
```