COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 16, 2013 9:00 A.M. – 1:00 P.M.

STOR 665 Questions (100 points in total)

1. (50 points) Bliss (1935) performed an experiment on adult flour beetles to assess whether the beetles had died after five hours of exposure to a different number of concentrations (doses) of gaseous carbon disulfide. The variables include the log10 dose of gaseous carbon disulfide in units of $\log_{10} CS_2 mg/l$, and whether or not the beetle died (1) or remained alive (0). The goal of the study is to understand the relationship between (log) dose and the probability of dying of a beetle.

   Some R analysis results are included below.

```
> beetles <- read.table("../datasets/beetles.txt", header=T)
> total <- table(beetles$log10.dose)
> alive.or.dead <- table(beetles$log10.dose, beetles$dead)
> prop.dead <- alive.or.dead[,2] / total
> data.sum<-cbind(alive.or.dead, total = total,  prop.dead)
> unique.log10.dose <- sort(unique(beetles$log10.dose))
> data.sum
        0  1 total prop.dead
1.6907 53  6    59 0.1016949
1.7242 47 13    60 0.2166667
1.7552 44 18    62 0.2903226
1.7842 28 28    56 0.5000000
1.8113 11 52    63 0.8253968
1.8369  6 53    59 0.8983051
1.861   1 61    62 0.9838710
1.8839  0 60    60 1.0000000
> beetles.m1 <- glm(dead ~ log10.dose, data=beetles, family=binomial)
> summary(beetles.m1)

Call:
glm(formula = dead ~ log10.dose, family = binomial, data = beetles)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4922  -0.5986   0.2058   0.4512   2.3820

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16 ***
log10.dose    34.270      2.912   11.77   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 645.44  on 480  degrees of freedom
Residual deviance: 372.47  on 479  degrees of freedom
AIC: 376.47

Number of Fisher Scoring iterations: 5

> summary(beetle.logit.log10dose)$cov.unscaled[1,2]
[1] -15.08189
```

(1a) (5 points) Explain the model fitted in the above R analysis. Derive the corresponding log likelihood function with clear notations.

(1b) (6 points) Show the distribution used is a member of the GLM exponential family, by writing the pdf in the canonical form. Identify the canonical parameter, θ_i, as well as the functions $a(\phi)$, $b(\theta_i)$, $c(y_i, \phi)$.

(1c) (11 points) Using the model `beetles.m1`, estimate the probability that a beetle will not survive with a log10 dose of 1.8. Produce a 95% CI for this probability.

```
  > beetles.m2 <- glm(prop.dead ~ unique.log10.dose, weights=total,
+                             data=beetles, family=binomial)
>
> summary(beetles.m2)

Call:
glm(formula = prop.dead ~ unique.log10.dose, family = binomial,
    data = beetles, weights = total)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5941  -0.3944   0.8329   1.2592   1.5940

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -60.717      5.181  -11.72   <2e-16 ***
unique.log10.dose   34.270      2.912   11.77   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom
AIC: 41.43
```

```
Number of Fisher Scoring iterations: 4
```

(1d) (10 points) Explain the model `beetles.m2` fitted in the above R analysis. Derive the corresponding log likelihood function with clear notations. Demonstrate the MLEs for the parameters (β_0, β_1) are the same under both likelihoods of models `beetles.m1` and `beetles.m2`.

(1e) (13 points) For the null models, explain why the deviances for models `beetles.m1` and `beetles.m2` are different. Show the detailed calculation for the null deviance of the model `beetles.m2`.

(1f) (5 points) Explain why the difference in deviances under the two models is the same.

2. (50 points) Consider the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, a;\ j = 1, \ldots, b, \tag{1}$$

where μ, β_j; $j = 1, \ldots, b$, are fixed but unknown, and α_i and ϵ_{ij} are independent random variables with mean 0 and $\text{var}(\alpha_i) = \sigma_\alpha^2$; $\text{var}(\epsilon_{ij}) = \sigma_\epsilon^2$.

(2a) (18 points) Suppose $\sigma_\alpha^2 = 0$. Consider the parameter $\theta = \sum_{j=1}^{b} c_j\beta_j$, where $c_j; j = 1, \ldots, b$, are constants.

(2a.1) (4 points) Give the condition that θ is a contrast.

(2a.2) (4 points) Explain when θ is estimable.

(2a.3) (10 points) Is the following statement true? If yes, please provide a detailed proof. If not, please provide a counter example.
"The parameter θ is a contrast if and only if θ is an estimable parameter."

(2b) (10 points) Write the model (1) in the matrix form; Calculate the covariance matrix of the response vector and its inverse matrix.

(2c) (10 points) Derive the BLUE for $\beta_j; j = 1, \ldots, b$.

(2d) (12 points) Derive the BLUP for $\alpha_i; i = 1, \ldots, a$.