PhD Prelim Exam
METHODS

Summer 2007
(Given on 7/10/07)

The questions in this problem are all based on a study of mastitis in dairy cows. Mastitis is an inflamation of the cow's udder that reduces milk production. It is a serious economic concern for dairy farmers. Three relevant features of mastitis are that:

it is more frequent at certain times of year,
it is more frequent in some cows than others, and
the effects of antibiotic treatments do not continue after the antibiotic is stopped.

Mastitis can be diagnosed either by a veterinary inspection or by measuring the sodium to potassium ratio (Na/K ratio) in the milk. Veterinary inspection has two possible outcomes: diseased or not diseased. Na/K ratio is a continuous outcome.

The investigators are comparing three treatments: a standard antibiotic treatment (STD), a new antibiotic treatment (NEW), and a placebo (P). They have 6 cows available to them. They have data from a preliminary study and a second more extensive study. Each of the four parts of this problem concerns different data arising from these two studies.

Part A. The preliminary study was a completely randomized design, i.e. the six cows were randomly assigned to treatments, with two cows per treatment. The questions in this part all concern the analysis of the Na/K ratio. Na/K means for the three treatments are:

Placebo	STD	NEW
10.0	8.50	8.64

1. Complete the following ANOVA table, compute the F statistic to test the null hypothesis of no difference between treatments, and approximate the p-value for that test (e.g. $0.001 < p < 0.01$, $p < 0.0001$, or some other range of values).

Source	d.f.	SS	MS	F	p-value
Treatment					
Error		7.5			
Corrected total		13.0			

2. The investigators are especially interested in the contrast between the Placebo and the average of the STD and NEW treatments. Test whether this difference = 0. Report your test statistic and an approximate p-value.

Part B. The investigators believe that the variability between cows and between times contributes to the large MSE in the preliminary study. Hence, they design a second study using a Latin Rectangle design with 6 cows and 9 periods in one barn. Each treatment is repeated three times for each cow and twice each period. Each period is 28 days long. For the questions in Part B there is one observation of the Na/K ratio for each combination of cow and period. The treatment allocation is:

	Period								
Cow	1	2	3	4	5	6	7	8	9
1	P	STD	NEW	STD	P	NEW	NEW	STD	P
2	STD	NEW	P	P	NEW	STD	STD	P	NEW
3	NEW	P	STD	NEW	STD	P	P	NEW	STD
4	STD	P	NEW	P	STD	NEW	STD	NEW	P
5	P	NEW	STD	STD	NEW	P	P	STD	NEW
6	NEW	STD	P	NEW	P	STD	NEW	P	STD

3. The investigators need an analysis that removes variability between cows and variability between periods from the error. Write out the statistical model and skeleton ANOVA table (sources of variability and associated d.f.) that is appropriate for this study. Make sure to define terms and subscripts in your model.

4. The desired inferences are narrow sense, i.e. specific to these cows and periods. Which terms in your model should be considered fixed and which should be considered random?

5. Identify the experimental unit and the observational unit in this study.

6. The Type III Sums-of-Squares can be obtained as a comparison between a full and a reduced model. Write out the appropriate full and reduced models used to calculate the Type III Sums-of-squares for treatments. Be sure to explain parameters and subscripts in your models.

7. The investigators are interested in the new antibiotic treatment because it has many fewer side effects than the standard treatment (data and analyses not shown here). Hence, if the two antibiotics have equivalent effects on the Na/K ratio, they will adopt the new one. Remember that a large p-value (e.g. 0.82) for the usual test of H0: $\mu_A - \mu_B = 0$ does not prove that $\mu_A = \mu_B$. One way to demonstrate equivalence is to reverse the null hypothesis, i.e. test $H0_{equiv}$: $\mu_A - \mu_B < \delta_l$ or $\mu_A - \mu_B > \delta_u$, where (δ_l, δ_u) is an a-priori chosen interval that is considered equivalent to 0. This can be tested using two one-sided tests. That is, reject $H0_{equiv}$ if you reject $H0_1$: $\mu_A - \mu_B < \delta_l$ and reject $H0_2$: $\mu_A - \mu_B > \delta_u$. If the test of $H0_1$ has a type I error rate of α and the test of $H0_2$ has a type I error rate of α, what can you say about the type I error rate for the test of $H0_{equiv}$? I.e., is the type I error equal to 0.05 or equal to some specific number other than 0.05, or can you only provide a range, > 0.05 or < 0.05? Explain your conclusion.

8. The treatment means are:

Placebo	STD	NEW
9.6	8.60	8.74

The MSE from the appropriate ANOVA is 0.025. The investigators choose the bounds δ_l = -0.85 and δ_u = 0.85. Test whether the STD and NEW antibiotics are equivalent using $\alpha = 0.05$ one-sided tests. Provide a one-sentence conclusion for the investigators.

Part C. The analyses in Part B use one observation of Na/K ratio per period. This is a subset of the full data set. The Na/K ratio was measured daily on the 10 days at the end of each of the 28-day treatment periods, so there are a total of 540 observations. One possible model for these data is:

$$Y_{ijk} = \mu_i + \nu_{ij} + \varepsilon_{ijk},$$

where μ_i represents the mean for a set of fixed effects (treatments and any relevant blocking variables) indexed by the subscript i, $\nu_{ij} \sim N(0, \sigma^2_\nu)$ is a random variable for each experimental unit indexed by i, j, and $\varepsilon_{ijk} \sim N(0, \sigma^2_e)$ is a random variable for each daily observation indexed by i, j, k. All random variables are independent. Included on the next page is part of SAS output from PROC MIXED. Certain parts have been deleted. Use this output to answer the next set of questions.

9. Test H0: $\sigma^2_\nu = 0$. Report the appropriate test statistic and appropriate p-value. If this is not possible from the included output, say what additional information you would need.

10. Estimate σ^2_ν and σ^2_e. If not possible from the included output, say what additional information you would need.

11. The study will be repeated in a different barn using the same (or a similar) Latin Rectangle design. Again, the goal is to estimate treatment means for the two antibiotics and the placebo treatment. It is time consuming to measure 540 milk samples, so the investigators would like to reduce the sample size. Which design would you recommend:
    A: 3 cows, 9 periods, 10 observations per experimental unit, or
    B: 6 cows, 9 periods, 5 observations per experimental unit?
    Explain your decision.

Part D. The other response collected in these studies is the outcome of the veterinary inspection. This is a yes/no response, observed once for each cow and period.

12. Here is the table of counts for each combination of treatment and response. These are summed over cows and periods.

Response	Placebo	STD	NEW
Yes	16	10	8
No	2	8	10

    Define π_i as the proportion of yes responses in treatment i. Assume each response is independent and test H0: $\pi_{\text{STD}} = \pi_{\text{NEW}}$.

13. The usual approach to the test in question 12 is one of many possible tests. Other tests are based on a comparison of two models. Give the full and reduced models that lead to another appropriate test of H0: $\pi_{\text{STD}} = \pi_{\text{NEW}}$. Again, assume responses are independent.

14. The models in question 13 assumes that responses are independent. This may be questionable because of variability among cows and among periods. Write down a full model that will allow you to estimate treatment differences (on an appropriate scale) after adjusting for variability among cows and among periods. In your parameterization, what function of the parameters represents the difference between treatments (again, the difference is measured on some appropriate scale).

The Mixed Procedure

Model Information

Data Set	WORK.COW
Dependent Variable	nak
Covariance Structure	Variance Components
Estimation Method	Type 3
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

<Some output deleted>

Number of Observations

Number of Observations Read	540
Number of Observations Used	540
Number of Observations Not Used	0

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square

<some output deleted>

trt	2	142.53	71.26	Var(Error) + 10 Var(e.u.) + Q(trt)
e.u.	38	11.59	0.305	Var(Error) + 10 Var(e.u.)
Error	486	1.27	0.00262	Var(Error)

<some output deleted>

Fit Statistics

-2 Res Log Likelihood	-1381.2
AIC (smaller is better)	-1377.2
AICC (smaller is better)	-1377.2
BIC (smaller is better)	-1373.2

1. Working from the numbers in the partial ANOVA table, you get:

Source	df	SS	MS	F	p value
Treatment	2	5.5	2.75	1.10	>0.10
Error	3	7.5	2.50		
c. Total	5	13.0			

   If you calculate SS_{trt} from the treatment means, you get a different value. This is because I simplified one part of the problem and forgot to adjust SS_{trt}.

2. The desired contrast is $\gamma = \mu_P - (\mu_{STD} + \mu_{NEW})/2$. The estimate $\hat{\gamma} = 10.0 - (8.5 + 8.64)/2 = 1.43$ with se = $\sqrt{2.5(1^2 + 0.5^2 + 0.5^2)/2} = 1.369$. The test statistic is $T = (\hat{\gamma} - 0)/se_{\hat{\gamma}} = 1.04$. This is compared to a t distribution with 3 d.f. The two-sided p-value is between 0.3 and 0.4 because $T_{3,0.80} = 0.978$ and $T_{3,0.85} = 1.250$.

3. $Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}$, $\varepsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
   where i indexes cows, j indexes periods and k indexes treatments. The model is an additive effects model with an overall constant, μ, cow effects, α_i, period effects, β_j, and treatment effects, τ_k. The errors are assumed independent normal with constant variance.

   The corresponding skeleton ANOVA table is:

Source	d.f.
Cow	5
Period	8
Treatment	2
error	38
c. Total	53

4. Cow, Period and Treatment are fixed; error is random.

5. e.u. = cow*period combination, o.u. = same

6. full: $Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}$, $\varepsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
   reduced: $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$, $\varepsilon_{ijk} \sim$ iid $N(0, \sigma^2)$

7. The test of $H0_{equiv}$ has a type I error rate $\leq \alpha$. To see this, consider the rejection regions for each of the one-sided tests. A picture helps considerably. $H0_l$ will reject if $T_l = (\overline{Y}_A - \overline{Y}_B - \delta_l)/s_d > T_{1-\alpha}$, where s_d is the se. of the difference and $T_{1-\alpha}$ is the $1 - \alpha$ quantile of the T distribution with the appropriate d.f. The null hypothesis is composite, so you need to consider where P[reject $H0_l$ | H0 true] is largest. This is at the boundary, $\mu_A - \mu_B = \delta_l$, where P[reject $H0_l$ | $H0_l$ true] = α. This rejection region for the test of $H0_l$ includes some values for which the second test (of $H0_u$) is not rejected. The desired rejection region, for $H0_{equiv}$, is the intersection of the two one-sided rejection regions, so P[reject $H0_{equiv}$ | $H0_{equiv}$ true] < P[reject $H0_{equiv}$ | $H0_{equiv}$ true]. Hence, the type I error rate for the equivalence test < α.

   Note: This is not a traditional multiple testing problem, for which the family-wise error would be > α. In the multiple testing problem, the family-wise error rate is the probability of rejecting one or more individual tests. The rejection region of the overall test is the **union** of rejection regions for each individual test. The equivalence test rejects Ho only if both individual tests are rejected. Its rejection region is the **intersection** of the individual tests.

8. Each treatment is used a total of 18 times, so the s.e. of the difference in means = $\sqrt{0.025 * 2/18} = 0.053$. The $\alpha = 0.05$ and $1 - \alpha = 0.95$ quantiles of the T distribution with 38 d.f. are approximately -1.686 and 1.686. The T statistic for the test of $H0_l = (8.74 - 8.60 - (-0.85))/0.053 = 18.7 > 1.686$, so you reject $H0_l$. The T statistic for the test of $H0_u = (8.74 - 8.60 - 0.85)/0.053 = -13.4 < -1.686$, so $H0_u$ is also rejected. Reject $H0_{equiv}$. Conclusion: The mean Na/K ratios for the Standard and New vaccines are similar. The null hypothesis of non-equivalence is rejected at $\alpha \leq 5\%$.

9. The desired test statistic is $F=MS_{eu}$ / MS_{error} = 116.7. This is compared to quantiles of the $F_{38,486}$ distribution. The p-value is $p < 0.0001$.

10. You have sufficient information for a method-of-moments, also called ANOVA, estimate of the variance component. $\hat{\sigma}^2_\nu = (0.305 - 0.0026)/10 = 0.030$.

11. You should recommend Design B. Reasons include:
    a) General principle of "replicate as high up as possible", i.e. use more cows not more observations per cow
    b) Estimate Var $\overline{Y}_{..k} = \sigma^2_\nu/\#eu + \sigma^2_e/\#obs$ for each design, using estimated variance components and the provided sample sizes. For design A (3 cows), each treatment occurs 9 times (3 times for each cow), so Var $\overline{Y}_{..k} = 0.030/9 + 0.0026/90 = 0.00336$. For design B (6 cows), each treatment occurs 18 times, so Var $\overline{Y}_{..k} = 0.030/18 + 0.0026/90 = 0.00170$. Design B gives a much smaller Var $\overline{Y}_{..k}$. Design B also has a smaller variance for any treatment difference or contrast because $\overline{Y}_{..P}$, $\overline{Y}_{..S}$, and $\overline{Y}_{..N}$ are independent conditional on the cow and period effects.

12. You have the information to do a Chi-square test for a 2 x 2 contingency table. The Placebo responses are ignored, since the hypothesis concerns only the STD and NEW treatments. Each cell has an expected count of 9, so χ^2 = 1/9 + 1/9 + 1/9 + 1/9 = 0.44. This is compared to a 1 d.f. Chi-square distribution. $p > 0.50$.

13. You could write the full and reduced models in various ways. Two reasonable possibilities are:
    a) In terms of Y_i, the number of Yes responses for each group
    full: $Y_i \sim$ Bin (N_i, π_i), where i indexes treatment (STD or NEW)
    reduced: $Y_i \sim$ Bin (N_i, π).
    b) In terms of individual observations, Y_{ijk},
    full: $Y_{ijk} \sim$ Bernoulli (π_i)
    reduced: $Y_{ijk} \sim$ Bernoulli (π)

14. It is easiest to generalize the model for individual observations

$$\begin{aligned} Y_{ijk} &\sim \text{Bernoulli } (\pi_{ijk}) \\ \text{logit } \pi_{ijk} &= \mu + \alpha_j + \beta_k + \tau_i \end{aligned}$$

where for consistency with the previous part, i indexes treatments and j and k index cows and periods. The difference between treatments is expressed as a log odds ratio, $\tau_{std} - \tau_{new}$.

Weight gain (in grams) of a group of animals is monitored at irregular intervals and recorded in days since birth over a period of 300 days from birth. Additional co-variates are animal sex (1 = male, 2 = female) and location of the animal (locations are G, Ha, Hb and *other*). A total of 770 records on 68 different animals are available.

1. In a first attempt, we will ignore dependencies among responses attributable to the fact that multiple measurements were made on animals.

   a) Set up a linear model for Log Weight depending on the other co-variates with main effects only. Use zero-sum constraints for the effects of qualitative factors. Write out the model in matrix form for the following five observations:

ID	Weight	Age	Sex	Location
102	124	133	2	Ha
53	450	205	2	G
88	174	136	1	other
105	233	158	1	Ha
138	957	246	2	Hb

   Make sure to explain all the notation you use and assumptions you make. We will refer to this model as Model A.

   b) Model A has a residual standard deviation of 0.2077, whereas a model that does not consider location as co-variate, has an increased standard deviation of 0.2094. Use these values to test whether location contributes to the model.

2. Now we will model dependencies among responses due to measurements on the same animal by including random effects.

   a) Write out a model generalizing model A by including random effects for each of the animals for intercept and age effect. We will call this model C.

   b) Illustrate the nature of the relationship between Age of an animal and the fitted values of model C with a hypothetical scatterplot.

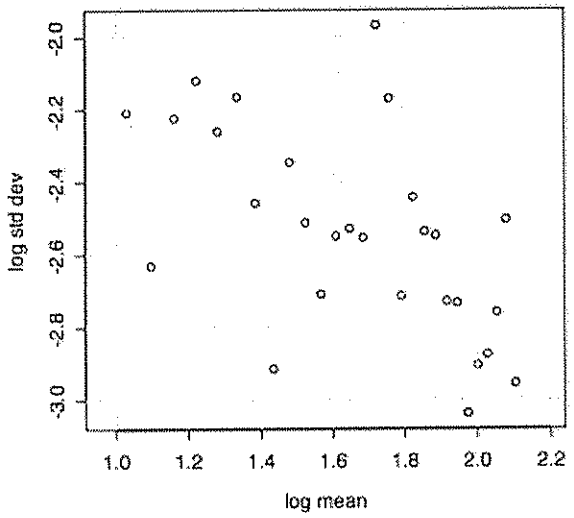
   c) You can find a fit of model C in the back, corresponding to `lmeC`. Based on this `R` report, determine the correlation between measurements on the same animal at 50 and 150 days after birth.

   d) Based on the output for model C in the back, determine, whether there is a significant difference in the log weight gain between locations 2 and 3 (Ha and Hb). You can assume a significance level of $\alpha = 0.05$.

   e) The following table contains maximized log-likelihood values for models A, C and some variations of model C. Discuss in detail, why these values allow or do not allow valid statistical comparisons. State (asymptotic) distributions and your conclusions. All values are restricted maximum likelihood (REML) values:

Model	REML
Model A	93.85
Model C	268.70
Model C, no random effect for age	259.00
Model C, no Location effect	275.50

f) The scatterplot below shows the log standard deviation of residuals versus log mean values of binned fitted values of model C. Interpret the graphic, explain its relevance to the model and describe in detail the next steps for modeling the data appropriately based on this evidence.



Selected R output

```
> lmA <- lm(log(Weight)~Age+Sex+Location, data=roo)
> summary(lmA)

Call:
lm(formula = log(Weight) ~ Age + Sex + Location, data = roo)

Residuals:
      Min        1Q    Median        3Q       Max
-0.705348 -0.136680  0.000354  0.142072  0.495638

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.0076506  0.0190225 105.541  < 2e-16 ***
Age          0.0206511  0.0001031 200.329  < 2e-16 ***
Sex1        -0.0322103  0.0076016  -4.237 2.54e-05 ***
Location1    0.0023657  0.0126191   0.187 0.851344
Location2   -0.0427737  0.0123483  -3.464 0.000562 ***
Location3    0.0007729  0.0140253   0.055 0.956067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2077 on 764 degrees of freedom
Multiple R-Squared: 0.9815,Adjusted R-squared: 0.9813
F-statistic:  8093 on 5 and 764 DF,  p-value: < 2.2e-16
>
> lmeC <- lmer(log(Weight)~Age+Sex+Location+(1+Age|ID), data=roo)
> summary(lmeC)
Linear mixed-effects model fit by REML
Formula: log(Weight) ~ Age + Sex + Location + (1 + Age | ID)
   Data: roo
    AIC    BIC logLik MLdeviance REMLdeviance
 -519.4 -477.6  268.7     -581.2       -537.4
Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 ID       (Intercept) 2.7139e-02 0.16474044
          Age         6.2358e-07 0.00078967 -0.520
 Residual             2.1169e-02 0.14549409
number of obs: 770, groups: ID, 68
```

```
Fixed effects:
             Estimate Std. Error t value
(Intercept)  1.9853012  0.0252602   78.59
Age          0.0207593  0.0001343  154.60
Sex1        -0.0167927  0.0185126   -0.91
Location1   -0.0090072  0.0325680   -0.28
Location2   -0.0367443  0.0309203   -1.19
Location3    0.0051695  0.0307316    0.17

Correlation of Fixed Effects:
          (Intr) Age    Sex1   Loctn1 Loctn2
Age       -0.646
Sex1      -0.045  0.006
Location1 -0.035  0.002 -0.101
Location2 -0.055 -0.015  0.051 -0.294
Location3 -0.089  0.019  0.085 -0.291 -0.210
>
```

Weight gain (in gram) of a group of animals is monitored in irregular intervals at days since birth over a period of 300 days. Additional co-variates are sex (1 = male, 2 = female) and location of the animal (locations are G, Ha, Hb and *other*). A total of 770 records on 68 different animals are available.

1. In a first attempt, we will ignore dependencies among responses due to records on the same animal and assume independence.

   a)

   A model of the form $Y = X\beta + \varepsilon$ *with* $\varepsilon \sim MVN(0, \sigma^2 I_{770\times 770})$ *for the above five observations can be written as*

$$\log \begin{pmatrix} 124 \\ 450 \\ 174 \\ 233 \\ 957 \end{pmatrix} = \begin{pmatrix} 1 & 133 & -1 & 0 & 1 & 0 \\ 1 & 205 & -1 & 1 & 0 & 0 \\ 1 & 136 & 1 & -1 & -1 & -1 \\ 1 & 158 & 1 & 0 & 1 & 0 \\ 1 & 246 & -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \varepsilon,$$

   where

   μ *is the overall mean of log Weight*
   α *is the expected log weight gain per day*
   β *is the expected effect of gender on log weight*
   γ_i *is the effect of location on log Weight, with* $i = 1, 2, 3, 4$ *and* $\sum \gamma_i = 0$

   b)

   The residual standard error s_A *of model A can be found as* $s_A^2 = SSE_A/(770 - 6)$, *i.e.* $SSE_A = s_A^2 \cdot 764 = 32.9584$ *and* $SSE_B = s_B^2 \cdot 767 = 33.6317$.
   Since B is included in A, we can set up an F test with

$$\frac{(SSE_B - SSE_A)/3}{SSE_A/767} = 5.22 \sim F_{3,767},$$

   which results in a significant p value, indicating that the factor is necessary.

2. Now we will regard dependencies among responses due to measurements on the same animal by including random effects.

   a)

$$y_i = \mu + \alpha \cdot age_i + \beta_{j(i)} + \gamma_{k(i)} + a_{\ell(i)} + b_{\ell(i)} age_i + \varepsilon_i$$

   where

μ *is the overall mean of log Weight*
α *is the expected log weight gain per day*
$\beta_{j(i)}$ *is the expected effect of gender on log weight*
$\gamma_{k(i)}$ *is the effect of location on log Weight, with* $k(i) = 1, 2, 3, 4$ *and* $\sum \gamma_{k(i)} = 0$
$a_{\ell(i)}$ *random effect for animal* $\ell(i)$ *with* $a \sim MVN(0, \sigma_a^2)$
$b_{\ell(i)}$ *random effect for the expected log weight gain for animal* $\ell(i)$ *with* $b \sim MVN(0, \sigma_b^2)$ *(random intercept)*
ε_i *with* $\varepsilon \sim MVN(0, \sigma^2 V)$

Random effects are assumed to be pairwise independent, i.e. $cov(\varepsilon, a) = cov(\varepsilon, b) = cov(a, b) = 0$.

b) Sketch the relationship between Age of an animal and the fitted values of model C in a scatterplot.

c)

The correlation between measurements on the same animal is

$$corr(y_i, y_j) = \frac{cov(y_i, y_j)}{\sqrt{var(y_i) var(y_j)}},$$

with

$$cov(y_i, y_j) = \sigma_a^2 + \sigma_b^2 \; age_i \; age_j$$

Then

$$corr(y_i, y_j) = \frac{0.027139 + 6 \cdot 10^{-7} \cdot 50 \cdot 150}{\sqrt{0.027139 + 6 \cdot 10^{-7} \cdot 50^2 + 0.021169} \cdot \sqrt{0.027139 + 6 \cdot 10^{-7} \cdot 150^2 + 0.021169}} = 0.570$$

d)

$$0.0419138 \pm t_{0.975,661} \cdot 0.1455 \cdot \sqrt{0.0023} = 0.0419138 \pm 0.01368$$

e)

Let's call model C without the random effect for age model C1 and the other variation model C2. Then comparisons based on the restricted maximum likelihood values between models A, C and C1 are all valid and lead to likelihood ratio tests. These test statistics have asymptotic χ^2 *distributions with degrees of freedoms based on the difference in number of parameters in the models. One complication is, that the asymptotic is violated because the null hypotheses of all of these tests are of the form* $\sigma_x^2 = 0$, *i.e. the parameter is on the boundary of its space. This leads to very conservative tests in this situation, since the actual asymptotic distribution is a mixture of a zero and a* χ^2 *distribution. However, the test results between all of the models are highly significant, indicating, that all of the random effects are necessary, and the conclusions are thereby not affected by the violation of the asymptotic behavior:*

Model A vs model C: $-2 \cdot (93.85 - 268.70) \sim \chi_2^2$ *undoubtedly significant*

Model C1 vs model C: $-2 \cdot (259.00 - 268.70) \sim \chi_1^2$, *again highly significant*

Comparisons to model C2 are not possible based on the restricted maximum likelihood values, since different fixed effects are involved, which changes the likelihood function involved.

f)

The downwards trend in the scatterplot indicates that (log) mean of the fitted values is not independent of the standard deviation in the residuals, which means that the error variance might better be modeled as a function of the mean, e.g. as:

$$y_i = \mu_i(\beta) + \sigma g(\mu_i(\beta), \theta)\varepsilon_i,$$

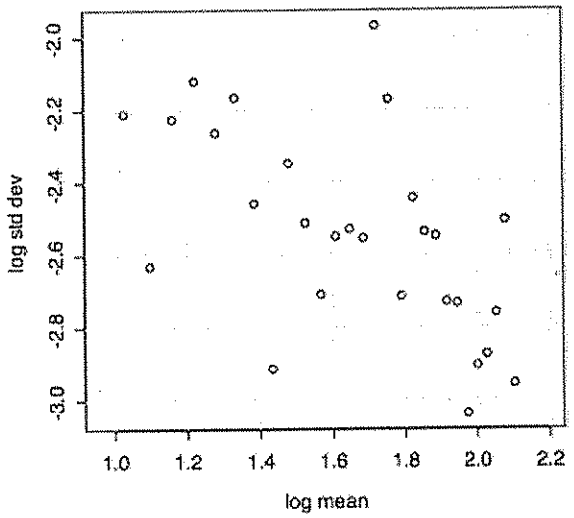
with $\varepsilon_i \sim N(0,1)$ *i.i.d.*
We could assume for $g(\mu_i(\beta), \theta)$ *a power of the mean transformation, i.e.*

$$g(\mu_i(\beta), \theta) = \mu_i(\beta)^\theta$$

Then the relationship between standard deviation and fitted values shown in the scatterplot is reflected in the model as

$$\log \sqrt{vary_i} = \log \sigma + \theta \log \mu_i(\beta)$$

and we get approximately constant variance, if we transform the response by taking it to the power of $1 - \theta$ *(Box-Cox transformation).*
From the scatterplot, we can get an eyeball estimate of θ *as -0.8, (a least square estimation of* θ *gives a value of -0.5026).*



Selected R output

```
> lmA <- lm(log(Weight)~Age+Sex+Location, data=roo)
> summary(lmA)

Call:
lm(formula = log(Weight) ~ Age + Sex + Location, data = roo)

Residuals:
      Min        1Q    Median        3Q       Max
-0.705348 -0.136680  0.000354  0.142072  0.495638

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.0076506  0.0190225 105.541  < 2e-16 ***
Age          0.0206511  0.0001031 200.329  < 2e-16 ***
Sex1        -0.0322103  0.0076016  -4.237 2.54e-05 ***
Location1    0.0023657  0.0126191   0.187 0.851344
Location2   -0.0427737  0.0123483  -3.464 0.000562 ***
Location3    0.0007729  0.0140253   0.055 0.956067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2077 on 764 degrees of freedom
Multiple R-Squared: 0.9815,Adjusted R-squared: 0.9813
F-statistic:  8093 on 5 and 764 DF,  p-value: < 2.2e-16
>
> lmeC <- lmer(log(Weight)~Age+Sex+Location+(1+Age|ID), data=roo)
> summary(lmeC)
Linear mixed-effects model fit by REML
Formula: log(Weight) ~ Age + Sex + Location + (1 + Age | ID)
   Data: roo
    AIC    BIC logLik MLdeviance REMLdeviance
 -519.4 -477.6  268.7     -581.2       -537.4
Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 ID       (Intercept) 2.7139e-02 0.16474044
          Age         6.2358e-07 0.00078967 -0.520
 Residual             2.1169e-02 0.14549409
number of obs: 770, groups: ID, 68

Fixed effects:
             Estimate Std. Error t value
(Intercept)  1.9853012  0.0252602   78.59
Age          0.0207593  0.0001343  154.60
Sex1        -0.0167927  0.0185126   -0.91
Location1   -0.0090072  0.0325680   -0.28
```

```
Location2   -0.0367443  0.0309203   -1.19
Location3    0.0051695  0.0307316    0.17

Correlation of Fixed Effects:
          (Intr) Age    Sex1   Loctn1 Loctn2
Age       -0.646
Sex1      -0.045  0.006
Location1 -0.035  0.002 -0.101
Location2 -0.055 -0.015  0.051 -0.294
Location3 -0.089  0.019  0.085 -0.291 -0.210
>
```

PhD Prelim Exam, Summer 2007

Methods III (601)

Note: The problem background given is rather extensive and includes many details. It is suggested you first read this without worrying about the details to get a "feel" for the problem, and then come back to get details that might be relevant to particular questions.

1 Problem Background

Short-term or acute toxicity tests are often conducted using the theory of *tolerance*, which stipulates that each individual organism has a certain tolerance to a potential toxicant. If exposed to any concentration of the toxicant less than its tolerance an organism will survive. If exposed to any concentration of the toxicant equal to or greater than its tolerance an organism will die. It is assumed that different organisms have different tolerances, and this then produces variability in the response of organisms to a given concentration of the toxicant.

The basic procedure used in acute toxicity tests is to expose groups of organisms to each of a number of concentrations of the toxicant for a fixed period of time. For example, 6 groups containing 20 organisms each might be exposed to 6 different concentrations of a toxicant, each exposure lasting for a period of 96 hours. At the end of the exposure period the number of organisms that have died (or lived) in each group is recorded.

A typical statistical model for the analysis of these types of studies is based on random variables defined as, for organism i exposed to concentration k of the toxicant,

$$X_{k,i} = \begin{cases} 0 & \text{if the organism does not die} \\ 1 & \text{if the organism does die} \end{cases}$$

The random variables $X_{k,i}$ are defined for $i = 1, \ldots, m_k$ and $k = 1, \ldots, K$, where m_k is the number of organisms exposed to concentration k and K is the number of different concentrations of the toxicant included in the study. Under an assumption that the response of organisms is independent, a random variable connected to the

proportion of organisms that die can be constructed for each exposure concentration as,

$$Y_k = \frac{1}{m_k} \sum_{i=1}^{m_k} X_{k,i}.$$

Notice here that the Y_k will correspond to observed proportions rather than observed counts.

A traditional generalized linear model can then be formulated by taking the Y_k to have binomial distributions with parameters p_k, which can be written in exponential dispersion family form as,

$$f(y_k|p_k) = \exp\left[m_k\left\{y_k\,\theta_k - b(\theta_k)\right\} + c(y_k, m_k)\right]; \quad y_k = 0, 1/m_k, \ldots, m_k/m_k, \qquad (1)$$

where

$$\begin{aligned} \theta_k &= \log\left(\frac{p_k}{1-p_k}\right), \\ b(\theta_k) &= -\log\left(\frac{1}{1+\exp(\theta_k)}\right), \\ c(y_k, m_k) &= \log\{m_k!\} - \log\{(m_k y_k)!\} - \log\{(m_k - m_k y_k)!\}. \end{aligned} \qquad (2)$$

Recall that under this formulation the first two central moments of Y_k are,

$$\begin{aligned} \mu_k \equiv E(Y_k) &= b'(\theta_k) = \exp(\theta_k)/\{1+\exp(\theta_k)\} = p_k \\ var(Y_k) &= (1/m_k)[\exp(\theta_k)/\{1+\exp(\theta_k)\}][1 - \exp(\theta_k)/\{1+\exp(\theta_k)\}] \\ &= (1/m_k)\mu_k(1-\mu_k). \end{aligned} \qquad (3)$$

A standard generalized linear model is then completed by specifying a link function $g(\cdot)$ such that,

$$g(\mu_k) = \beta_0 + \beta_1 x_k, \qquad (4)$$

where x_k is some measure of the concentration to which group k was exposed (e.g., concentration in original units or log units).

The connection between the standard generalized linear model of expressions (1) through (4) and the theory of tolerance described at the beginning of this question is that there is a correspondence between link function $g(\cdot)$ and the distribution of tolerances in a "population" of organisms. For example, a logit link $g(\mu) = \log\{(\mu)/(1-\mu)\}$ implies a logistic distribution of tolerances in the population, a

probit link $g(\mu) = \Phi^{-1}(\mu)$ for $\Phi(\cdot)$ the cdf of a normal distribution implies a normal distribution of tolerances, and a complementary log-log link $g(\mu) = \log\{-\log(1-\mu)\}$ implies an extreme value distribution of tolerances.

In general, the relation between a link function $g(\cdot)$ and the distribution of tolerances can be summarized as follows.

- Let $T_{k,i}$ denote a random variable connected with the tolerance of organism i to be exposed to concentration k of the toxicant.
- Let $G(\cdot)$ denote the common distribution of the $T_{k,i}$; $i = 1, \ldots, m_k$; $k = 1, \ldots, K$, and let μ_T and σ_T^2 denote location and scale parameters for this distribution. Thus, we assume that G defines a location-scale family of distributions, although note that location parameters may represent the mode rather than the expected value of a distributon.
- Standardize tolerances as $Z_{k,i} = (T_{k,i} - \mu_T)/\sigma_T$, and then assume $Z_{k,i} \sim iidG(z|0,1)$ where 0 is in the position of the location parameter and 1 is in the position of the scale parameter.
- If $x_1, \ldots, x_k$ are the concentrations to which groups of organisms are exposed, and $\delta_k = (x_k - \mu_T)/\sigma_T$, then $\mu_k = G(\delta_k|0,1)$, and expression (4) becomes,

$$g(\mu_k) = G^{-1}(\mu_k|0,1) = \frac{x_k - \mu_T}{\sigma_T} = -\frac{\mu_T}{\sigma_T} + \frac{1}{\sigma_T} x_k = \beta_0 + \beta_1 x_k. \qquad (5)$$

The goal is then to estimate the parameters of $T_{k,i} \sim G(t|\mu_T, \sigma_T)$ through $\beta_0 = -\mu_T/\sigma_T$ and $\beta_1 = 1/\sigma_T$.

Because there is a correspondence between the link function of the generalized linear model $g(\mu_k)$ and the cdf of the tolerance distribution G, it has been suggested that one might employ a parameterized link function which then defines an entire class of tolerance distributions indexed by the link function parameter. Specifically, one such link family is,

$$g(\mu|\lambda) = \log\left[\frac{(1-\mu)^{-\lambda} - 1}{\lambda}\right]; \quad \lambda > 0. \qquad (6)$$

This family of functions includes the logit link for $\lambda = 1$ and the complementary log-log link in the limit as $\lambda \to 0$. Figure 1 presents the response functions that

correspond to the link function (6) using a number of values of λ ranging from 0.005 to 3.5. Each curve was produced from the parameter values $\beta_0 = -5$ and $\beta_1 = 1$. Smaller values of λ correspond to curves that rise more rapidly while larger values of λ produce curves that rise more slowly.

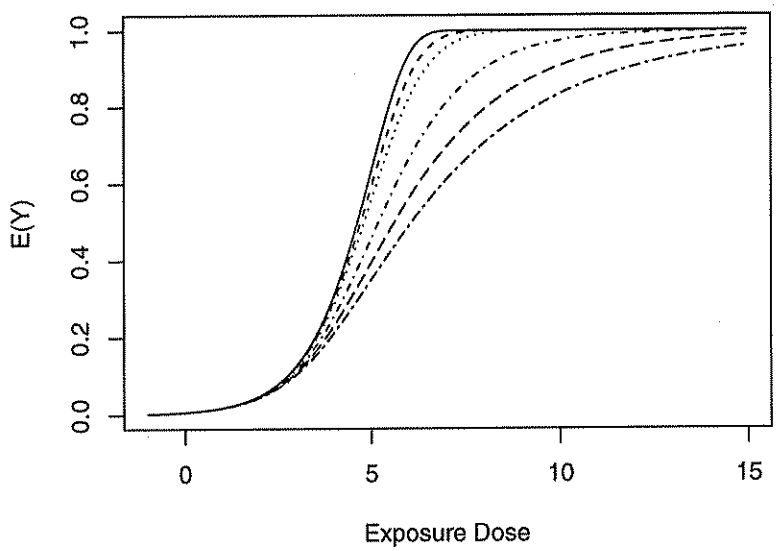


Figure 1: Response curves for a variety of λ.

Using the relation $g(\mu|\lambda) = G_\lambda^{-1}(\mu|0,1)$ we have that the cumulative density function of standardized tolerances is,

$$G_\lambda(t|0,1) = 1 - \frac{1}{\{\lambda \exp(t) + 1\}^{1/\lambda}}.$$

With regression coefficients $\beta_0 = -\mu_T/\sigma_T$ and $\beta_1 = 1/\sigma_T$, a location-scale transformation produces the tolerance distribution for the $T_{k,i};\ i = 1, \ldots, m_k;\ k = 1, \ldots, K$ as,

$$G_\lambda(t|\mu_T, \sigma_T) = 1 - \frac{1}{\{\lambda \exp\{(t-\mu_T)/\sigma_T\} + 1\}^{1/\lambda}}, \tag{7}$$

which has corresponding density function,

$$g_\lambda(t|\mu_T, \sigma_T) = \frac{\exp\{(t-\mu_T)/\sigma_T\}}{\sigma_T \left[\lambda \exp\{(t-\mu_T)/\sigma_T\} + 1\right]^{1+1/\lambda}}; \quad -\infty < t < \infty. \tag{8}$$

A number of densities of the form of expression (8) are presented in Figure 2. All of the densities graphed in Figure 2 have $\mu_T = 5$ and $\sigma_T = 1$, but each curve corresponds to a different value of λ. The values of λ used in Figure 2 range from 0.005 to 3.5; smaller λs give higher density at the mode and less right tail while larger λ give lower density at the mode and heavier right tails.

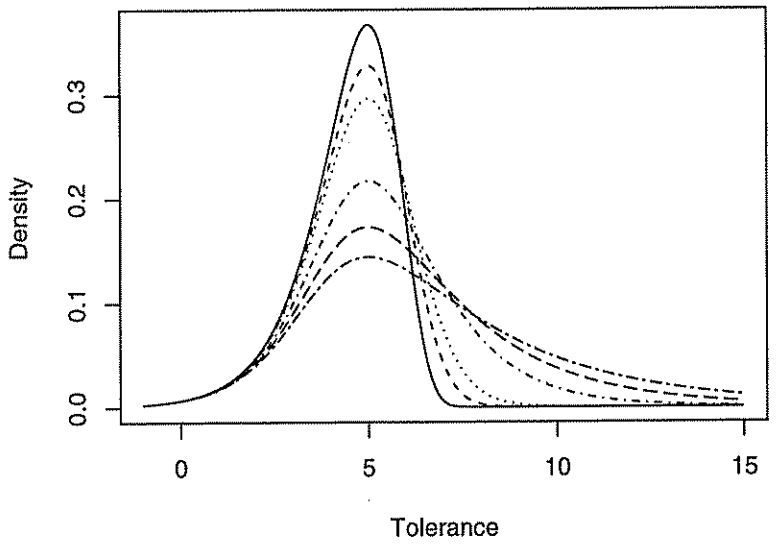


Figure 2: Tolerance distributions for a variety of λ.

2 Questions

1. Four statisticians we will call Statisticians 1, 2, 3, and 4, are contemplating using the family of link functions (6) to analyze data from a short-term toxicity test through application of a generalized linear model with binomial random component (1). The question of concern is how to deal with the link function parameter λ in estimation and inference.

   A basic division has emerged among our statisticians, with Statistician 1 taking the position that choosing a value for λ should be considered a part of model

formulation. He suggests fitting models with 3 or maybe 4 values of λ held fixed for estimation of the other parameters and then selecting an appropriate value based on model diagnostics such as residual plots. The remaining statisticians (2, 3, and 4) all hold that λ should be estimated as an additional parameter in the model.

(a) Give a short argument supporting the position of Statistician 1.

(b) Give a short argument supporting the position of the other three statisticians.

2. The position of statisticians 2, 3, and 4 wins the day (because I said so, that's why!) but there is not complete agreement among the individuals in this group as to how to approach estimation of λ. Statistician 2 wants to estimate λ using maximum likelihood and also wants to estimate the uncertainty associated with this value, but is not concerned with taking uncertainty in the estimate of λ into account when estimating uncertainty in the other parameters. Describe a procedure that would meet the needs of Statistician 2. Define quantities that would be involved in this procedure and indicate what should be done with those quantities.

3. Statistician 3 wants everything attained by Statistician 2, with the addition that uncertainty in estimation of λ is allowed to affect uncertainty about estimates of the other parameters. This statistician notes that estimation and inference for standard generalized linear models is easily accomplished through maximum likelihood and regular asymptotic results. In particular, she notes that this model contains no dispersion parameter that needs estimation, and that the likelihood equations for a generalized linear model with a **fixed link** function can be written as follows. Let $L_k = \log\{f(y_k|\theta_k)\}$, and let $g(\mu_k) = \beta_0 + \beta_1 x_k = \eta_k$. Then,

$$\frac{\partial}{\partial \beta_j} L_k(\boldsymbol{\beta}) = \frac{\partial L_k}{\partial \theta_k} \frac{d\theta_k}{d\mu_k} \frac{d\mu_k}{d\eta_k} \frac{\partial \eta_k}{\partial \beta_j}; \quad j = 0, 1, \tag{9}$$

and

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{k=1}^{K} \frac{\partial}{\partial \beta_j} L_k(\boldsymbol{\beta}).$$

She also notes that, using the probability mass function (1) and its exponential family properties in (2) and (3),

$$\begin{aligned}
\frac{\partial L_k}{\partial \theta_k} &= m_k\{y_k - b'(\theta_k)\} = m_k(y_k - \mu_k), \\
\frac{d\theta_k}{d\mu_k} &= \left(\frac{d\mu_k}{d\theta_k}\right)^{-1} = \frac{1}{b''(\theta_k)} = V^{-1}(\mu_k), \\
\frac{d\mu_k}{d\eta_k} &= \left(\frac{d\eta_k}{d\mu_k}\right)^{-1} = \left(\frac{dg(\mu_k)}{d\mu_k}\right)^{-1}, \\
\frac{\partial \eta_k}{\partial \beta_j} &= \begin{cases} 1 & j = 0 \\ x_k & j = 1 \end{cases}
\end{aligned}$$

Then, defining

$$W_k = \left[\left(\frac{d\eta_k}{d\mu_k}\right)^2 V(\mu_k)\right]^{-1},$$

we arrive at,

$$\begin{aligned}
\frac{\partial L_k}{\partial \beta_0} &= m_k(y_k - \mu_k)W_k\frac{d\eta_k}{d\mu_k} \\
\frac{\partial L_k}{\partial \beta_1} &= m_k(y_k - \mu_k)W_k\frac{d\eta_k}{d\mu_k}x_k.
\end{aligned} \tag{10}$$

Give modifications to expressions (9) and (10) that will allow Statistician 3 to accomplish her goals. Be careful in the use of d and ∂.

4. Statistician 4 is the Bayesian in the crowd, but he needs some help. What are the parameters in this model for which we would need to specify prior distributions, and what might reasonable prior distributions be (give names of distributions)? An overall Gibbs sampling algorithm would be one choice for simulating from the posterior for this model. Give a general form for distributions that would be needed, and describe how one might sample values from these distributions (e.g., built-in R functions, or the name of particular techniques for basic simulation). *Note: Do not get into technical derivations for specific distributions. Try to use generic distributional notation to describe a procedure.*

5. **Note: see the example answer following this question to get a better idea of what type of answer is being looked for.**

   Assuming that λ is a fixed parameter seems entirely appropriate for a single toxicity test conducted with organisms selected to have essentially the same characteristics (e.g., age, condition). But different "cohorts" (groups of some type) of organisms might well exhibit different tolerances to a given toxicant. Consider a toxicity test conducted using a moderate to large number of different cohorts (say 20 to 30 cohorts) with each cohort exposed to a range of concentrations. This could be thought of as a moderate to large number of individual toxicity tests of the kind described previously, in which we assume separate populations or sub-populations of organisms which might each have their own tolerance distribution.

   If we have no good understanding of how (or perhaps even why) various "cohorts" of organisms might differ in tolerance, formulate a model that could be used to analyze the entire set of toxicity tests, **assuming that the regression parameters β_0 and β_1 are the same for all cohorts**. Specify all parts of the model starting from scratch. Write down the entire model explicitly. You may use generic notation for density or mass functions, that is, use $p(x)$ to denote the density or mass function of a random variable X, $p(x, y)$ for the joint of two random variables X and Y or $p(x|y)$ for the conditional of X given Y. For the model you formulate, write out a general form for the likelihood or log likelihood. Indicate how you might accomplish the necessary computations for maximum likelihood estimation, and describe an approach you might use to produce inferential quantities (e.g., standard errors or interval estimates).

Example Answer on Question 5

To give you a better idea of what is being asked for in question 5, the following is what might suffice for a nonlinear model with random effects and additive errors (which is, of course, not the model you should be using):

Let $Y_{k,j}$ be random variables associated with the response for individual j in

group k.

Let $x_{k,j}$ be a covariate associated with individual j in group k.

Assume the model structure, for $\beta_k > 0$,:

$$Y_{k,j} = g(x_{k,j}, \beta_k) + \sigma\epsilon_{k,j}; \quad j = 1, \ldots, n_k; \; k = 1, \ldots K,$$

where $g(\cdot)$ is a known smooth function, $\epsilon_{k,j}$ are independent and identically distributed random variables following a location-scale distribution with $E(\epsilon_{k,j}) = 0$ and $var(\epsilon_{k,j}) = 1$, and β_k are independent and identically distributed random variables having a distribution with density $p(\beta_k|\theta)$ for some fixed parameter $\theta \in \Theta$. To compute maximum likelihood estimates of θ and σ we would need to evaluate a log likelihood of the following form.

$$\begin{aligned} L(\theta, \sigma) &= \sum_{k=1}^{K} log\{p(y_{k,1}, \ldots, y_{k,n_k}|\theta, \sigma)\} \\ &= \sum_{k=1}^{K} \log\left\{\int_0^\infty p(y_{k,1}, \ldots, y_{k,n_k}|\beta_k, \sigma)\, p(\beta_k|\theta)\, d\beta_k\right\} \\ &= \sum_{k=1}^{K} \log\left\{\int_0^\infty \prod_{j=1}^{n_k} p(y_{k,j}|\beta_k, \sigma)\, p(\beta_k|\theta)\, d\beta_k\right\}. \end{aligned}$$

Here, if $p_e(\epsilon|\sigma)$ is the density of the *iid* error terms $\epsilon_{k,j}; \; j = 1, \ldots, n_k; \; k = 1, \ldots, K$, then

$$p(y_{k,j}|\beta_k, \theta) = \frac{1}{\sigma} p_e\left(\frac{y_{k,j} - g(x_{k,j}, \beta_k)}{\sigma}\right).$$

If no closed form is available for these integrals they could be evaluated using numerical integration such as Gaussian quadrature (as performed by the R function *integrate*). If it is desired to use first and/or second derivatives, they could also be evaluated using numerical integration, assuming that we can interchange limits by passing derivatives under the integral. If second derivatives are available, intervals could be produced using standard errors estimated as diagonal elements of the inverse observed information matrix.

There are several ways to correctly answer Question 5. What is important is clarity of expression and notation. Make certain all of your expressions are consistent with one another.

End of Example Answer on Questions 5.

6. Statistician 1 (from question 1) remains unconvinced that estimation of parameters in link functions is a good idea. He suggests an alternative to the your model from Question 5 in which a common fixed link function is chosen for all cohorts of organisms, but the regression parameters are allowed to be random variables that differ in value among cohorts. Statistician 4 (the Bayesian) also likes this model because it fits into Winbugs easily. After discussing this possibility, Statisticians 1 and 4 question whether there is even really a difference in the way that this new model and your model from Question 5 are able to represent the problem. Describe a procedure or procedures (totally or at least primarily in words) by which you might examine this issue of whether your model of Question 5 and this alternative proposed by Statisticians 1 and 4 are truly representing the problem in different ways.

7. For this question, consider again the simpler case of one toxicity test with K groups of organisms as in Questions 1 and 2 (i.e., no differing cohorts of organisms). To this point, we have assumed that in the family of link functions of expression (6) we had $\lambda > 0$, and that the support of the corresponding tolerance density in expression (8) is $-\infty < t < \infty$. But there is no reason in (6) that λ must be strictly positive. Negative values of λ lead to perfectly valid link functions that produce values of $g(\mu_k|\lambda)$ restricted to (0, 1), which *is* necessary for a valid model. Interestingly, however, if $\lambda < 0$ in (6), then the support of (8) becomes $t < \mu_T + \sigma_T \log(-1/\lambda)$. Relative to conducting inference using Wald theory (asymptotic normality of maximum likelihood estimates with variances as given by inverse information) explain why this might or might not be cause for concern.

   Hint: Consider various inferential statements you might wish to make such as interval estimation of regression model parameters, or interval estimation of quantiles of the tolerance distribution.

PhD Preliminary Examination – 2007

Answers – Methods Question 3

These are a sketch of the answers hoped for. Other possibilities might exist for some of the questions that would be entirely adequate if they are both technically correct and logically consistent.

1. (a) An argument for treating selection of λ as a part of model formulation would be that λ essentially determines the shape (or set of shapes) that are possible for the expectation function of the generalized linear model. We typically specify a fixed form for regression functions. In addition, it appears unlikely (from Figure 1) that small differences in λ will produce much of a difference in the response function, making selection of λ somewhat analogous to selection of a power to be used in the family of power transformations (in a problem for which we might want to do that). Thus, it seems entirely reasonable to fit models with a small number of fixed link functions (i.e., values of λ) and choose the one that seems to best fit the data.

   (b) An argument for considering λ as a parameter in need of estimation is that determination of the tolerance distribution is a central question of scientific importance. The correspondence between values of λ and tolerance distributions means that by estimating λ we actually estimate the tolerance distribution. Thus, estimation of λ brings a central portion of the scientific question of interest into the model in a way that allows us to make inference about it.

2. An approach based on profile likelihoods would satisfy the needs of statistician

   2. To this end, define the log likelihood as

$$L(\lambda, \beta_0, \beta_1) \equiv \sum_{k=1}^{K} \log\{f(y_k|\theta_k)\},$$

where f is given in expression (1) of the exam question and, using the notation of that expression, $\mu_k = b'(\theta_k)$ and $g(\mu_k|\lambda) = \beta_0 + \beta_1 x_k$. Then the appropriate profile log likelihood for estimation of λ is,

$$L^p(\lambda) = \max_{\beta_0,\beta_1} L(\lambda, \beta_0, \beta_1). \tag{1}$$

Let Λ denote the parameter space of λ. The maximum likelihood estimate of λ is then that value $\hat{\lambda} \in \Lambda$ such that,

$$L^p(\hat{\lambda}) \geq L^p(\lambda); \quad \lambda \in \Lambda.$$

A picture of an idealized result from this procedure would graph $L_n^p(\lambda)$ against λ as in Figure 1 below. The scale of the ordinate would depend on whether or not constants were included or dropped in computation of $L(\lambda, \beta_0, \beta_1)$. The profile log likelihood of expression (1) could be maximized using a one-dimensional search algorithm (e.g., golden search) or simply through computation on a very fine one-dimensional grid such as might be used to produce a picture like Figure 1. Once the maximum likelihood estimate $\hat{\lambda}$ is obtained, an interval estimate can be produced through the same approach used with normed profile likelihoods. This is because, at $\hat{\lambda}$, the maximum likelihood estimates of β_0 and β_1 will also be available by definition of $L^p(\lambda)$ in the first place; this is true because the full parameter vector has been partitioned into two pieces. Specifically, a $(1-\alpha)100\%$ interval estimate of λ would be,

$$\{\lambda : -2[L^p(\lambda) - L^p(\hat{\lambda})] \leq \chi^2_{1,1-\alpha}\}.$$

Note: This is a more extensive answer than students would be expected to produce in a prelim exam. Any of the material presented here, however, could be relevant to a shorter and less complete answer.

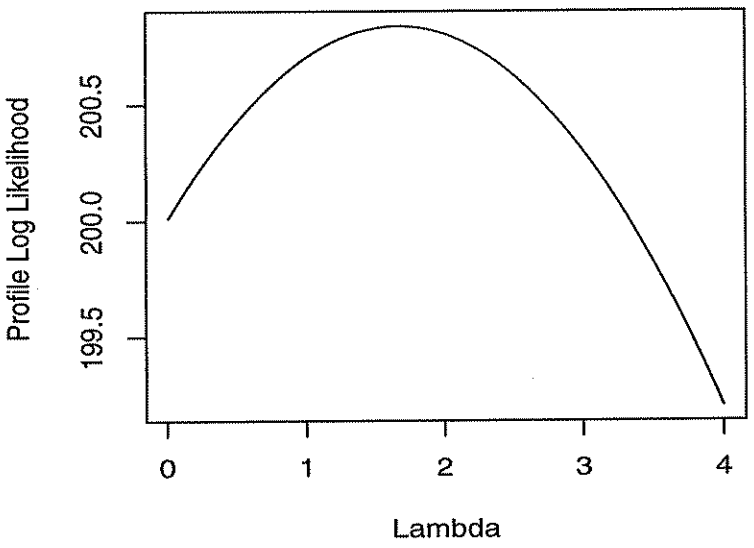


Figure 1: Idealized Profile Likelihood for Estimation of λ.

3. The goals of statistician 3 can be accomplished through evaluation of the information matrix, either expected or observed versions. While this could be done simply by computing second derivatives of the log likelihood at the maximum likelihood estimate found as in the answer to question 2, one could also accomplish simultaneous maximum likelihood estimation by modifying expressions (9) and (10) of the exam question in the following way. Rather than,

$$\frac{\partial}{\partial \beta_j} L_k(\boldsymbol{\beta}) = \frac{\partial L_k}{\partial \theta_k} \frac{d\theta_k}{d\mu_k} \frac{d\mu_k}{d\eta_k} \frac{\partial \eta_k}{\partial \beta_j},$$

which is expression (9) from the exam question, we would use,

$$\frac{\partial}{\partial \beta_j} L_k(\boldsymbol{\beta}, \lambda) = \frac{\partial L_k}{\partial \theta_k} \frac{d\theta_k}{d\mu_k} \frac{\partial \mu_k}{\partial \eta_k} \frac{\partial \eta_k}{\partial \beta_j}$$

and

$$\frac{\partial}{\partial\lambda}L_k(\boldsymbol{\beta},\lambda) \;=\; \frac{\partial L_k}{\partial\theta_k}\,\frac{d\theta_k}{d\mu_k}\,\frac{\partial\mu_k}{\partial\lambda} \tag{2}$$

For modification of (10) from the exam question, note that two things have changed in modification of expression (9). First, $d\mu/d\eta$ has been replaced with $\partial\mu/\partial\eta$ in the derivatives with respect to β_j; $j = 0, 1$, and second the additional factor of $\partial\mu/\partial\lambda$ appears in the derivative with respect to λ.

Now, μ_k may be defined as an implicit function of λ_k and η_k as $g(\mu_k|\lambda) - \eta_k = 0$, so that the implicit function gives,

$$\begin{aligned}
\frac{\partial\mu_k}{\partial\eta_k} &= -\frac{-1}{\dfrac{\partial g(\mu_k|\lambda)}{\partial\mu_k}} = \left(\frac{\partial\eta_k}{\partial\mu_k}\right)^{-1}, \\
\frac{\partial\mu_k}{\partial\lambda} &= -\frac{\dfrac{\partial g(\mu_k|\lambda)}{\partial\lambda}}{\dfrac{\partial g(\mu_k|\lambda)}{\partial\mu_k}} = -\frac{\partial\eta_k}{\partial\lambda}\left(\frac{\partial\eta_k}{\partial\mu_k}\right)^{-1}.
\end{aligned} \tag{3}$$

With W_k as defined in the question, this leads to the modification of (10) as,

$$\begin{aligned}
\frac{\partial L_k}{\partial\beta_0} &= m_k(y_k - \mu_k)W_k\frac{\partial\eta_k}{\partial\mu_k} \\
\frac{\partial L_k}{\partial\beta_1} &= m_k(y_k - \mu_k)W_k\frac{\partial\eta_k}{\partial\mu_k}x_k \\
\frac{\partial L_k}{\partial\lambda} &= m_k(y_k - \mu_k)W_k\frac{\partial\eta_k}{\partial\mu_k}\left(-\frac{\partial\eta_k}{\partial\lambda}\right).
\end{aligned} \tag{4}$$

4. The parameters requiring priors in this model are β_0, β_1, and λ. What is necessary is formulation of a joint prior, $\pi(\beta_0, \beta_1, \lambda)$. While any number of possible suggestions might be reasonable for specification of this prior, one easy solution would be to formulate the prior as a product of individual components,

$$\pi(\beta_0, \beta_1, \lambda) = \pi(\beta_0)\,\pi(\beta_1)\,\pi(\lambda).$$

Given the family of link functions in expression (6) of the question, one might consider

$$\pi(\beta_0) \;=\; N(B_0, \tau_0^2)$$

$$\begin{aligned} \pi(\beta_1) &= N(B_1, \tau_1^2) \\ \pi(\lambda) &= Gamma(\alpha, \beta) \end{aligned} \tag{5}$$

Values for B_0 and B_1 could perhaps be chosen from previous toxicity tests, τ_0^2 and τ_1^2 might be taken rather large, and α and β to produce a small expected value (e.g., 0.5) and large variance (e.g., 1000).

With any choice of priors estimation would most likely need to be conducted using Markov Chain Monte Carlo methods. A basic Gibbs algorithm would seem appropriate here, and full conditionals would have the simple forms,

$$\begin{aligned} p(\beta_0|\beta_1, \lambda, \boldsymbol{y}) &\propto f(\boldsymbol{y}|\beta_0, \beta_1, \lambda)\, \pi(\beta_0) \\ p(\beta_1|\beta_0, \lambda, \boldsymbol{y}) &\propto f(\boldsymbol{y}|\beta_0, \beta_1, \lambda)\, \pi(\beta_1) \\ p(\lambda|\beta_0, \beta_1, \boldsymbol{y}) &\propto f(\boldsymbol{y}|\beta_0, \beta_1, \lambda)\, \pi(\lambda). \end{aligned} \tag{6}$$

Here, sampling would most likely be approached using a rejection or adaptive rejection algorithm, although a standard Metropolis-Hastings algorithm could also be tried.

Although improper priors could be chosen for any or all of the parameters, one reason we might want to avoid this would be the need to demonstrate that the resultant posteriors are proper. Given the transformations necessary to move from parameters (β_0, β_1 and λ) to μ_k to θ_k in the likelihood, this could prove to be a challenge. Choosing all priors to be proper avoids this potential complication, although sensitivity of results to the choice of priors would then be more of an issue. If estimation with a given set of priors is quite difficult this might in itself be a substantial difficulty.

5. To formulate an appropriate model,

Let $Y_{g,k}$ denote random variables associated with the number of organisms in

cohort g and concentration k that die, for $g = 1, \ldots, G$ and $k = 1, \ldots, K_g$.

Let $m_{g,k}$ denote the number of organisms from cohort g exposed to concentration k; $g = 1, \ldots, G$; $k = 1, \ldots, K_g$.

Let $x_{g,k}$; $k = 1, \ldots, K_g$ denote the concentrations of the toxicant to which cohort g is exposed, for $g = 1, \ldots, G$.

Let β_0 and β_1 be fixed unknown parameter values. Assume that, for a given value of λ_g, the $Y_{g,k}$; $k = 1, \ldots, K_g$ are conditionally independent with probability mass functions,

$$f(y_{g,k}|\lambda_g) = \exp\left[m_{k,g}\{y_{g,k}\theta_{g,k} - b(\theta_{g,k}\} + c(y_{g,k}, m_{g,k})\right]; \quad y_{g,k} = 0, 1, \ldots, m_{g,k},$$

where $\theta \in (-\infty, \infty)$, and

$$\begin{aligned} b(\theta_{g,k}) &= -\log\left(\frac{1}{1+\exp(\theta_{g,k})}\right) \\ c(y_{g,k}, m_{g,k}) &= \log\{m_{g,k}!\} - \log\{(m_{g,k}y_{g,k})!\} - \log\{(m_{g,k} - m_{g,k}y_{g,k})!\}. \end{aligned}$$

Note that $\mu_{g,k} \equiv E(Y_{g,k}) = b'(\theta_{g,k})$ and $var(Y_{g,k}) = (1/m_{g,k})\mu_{g,k}(1-\mu_{g,k})$.

Assume that $g(\mu_{g,k}|\lambda_g) = \beta_0 + \beta_1 x_{g,k}$ where,

$$g(\mu_{g,k}|\lambda_g) = \log\left[\frac{(1-\mu_{g,k})^{-\lambda_g} - 1}{\lambda_g}\right]; \quad \lambda_g > 0.$$

Then, let the $\{\lambda_g : g = 1, \ldots, G\}$ be independent random variables having a common probability density function $h(\lambda_g|\psi)$; $\lambda_g > 0$ for some fixed parameter $\psi \in \Psi$.

To develop the likelihood or log likelihood, let $\boldsymbol{y}_g \equiv (y_{g,1}, \ldots, y_{g,K_g})^T$ for $g = 1, \ldots, G$, let $\boldsymbol{y} \equiv (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_G^T)^T$, and let $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_G)^T$. Note that,

$$\begin{aligned} p(\boldsymbol{y}_g|\beta_0, \beta_1, \lambda_g) &= \prod_{k=1}^{K_g} f(y_{g,k}|\beta_0, \beta_1, \lambda_g) \\ p(\boldsymbol{y}|\beta_0, \beta_1, \boldsymbol{\lambda}) &= \prod_{g=1}^{G} p(\boldsymbol{y}_g|\beta_0, \beta_1, \lambda_g). \end{aligned} \tag{7}$$

The first line of (7) follows from conditional independence of the $Y_{g,k}$ given λ_g and the second line follows from independence of the λ_g. The marginal likelihood is then,

$$\begin{aligned} p(\boldsymbol{y}|\beta_0,\beta_1,\psi) &= \int_0^\infty \cdots, \int_0^\infty f(\boldsymbol{y}|\beta_0,\beta_1,\boldsymbol{\lambda}) \prod_{g=1}^{G} h(\lambda_g|\psi)\, d\boldsymbol{\lambda} \\ &= \prod_{g=1}^{G} \int_0^\infty f(\boldsymbol{y}_g|\beta_0,\beta_1,\lambda_g)\, h(\lambda_g|\psi)\, d\lambda_g. \end{aligned} \tag{8}$$

The last line of (8) is useful for derivation of the log likelihood in terms of one-dimensional integrals as,

$$L(\beta_0,\beta_1,\psi) = \sum_{g=1}^{G} \log\left[\int_0^\infty f(\boldsymbol{y}_g|\beta_0,\beta_1,\lambda_g)\, h(\lambda_g|\psi)\, d\lambda_g\right]. \tag{9}$$

It is not likely that a density $h(\lambda_g|\psi)$ can be found that combines nicely in a mathematical way with the data model. The marginal log likelihood in (9), however, now involves only one-dimensional integrals. The log likelihood can be evaluated numerically without great difficulty. Assuming that derivatives can be passed under the integral this may also be true for at least first derivatives (and possibly even second) allowing maximum likelihood estimates to be found using gradient or Newton-type algorithms. If derivatives prove too unwieldy, direct search algorithms should prove adequate as the parameter space is only three-dimensional. Intervals could be produced by parametric bootstrap or possibly using normed profile likelihoods (if covariances between estimates are not desired).

6. Determining whether there is indeed a difference in the way the two models represent the situation could be approached through computation of conditional and marginal model moments, or simulation of data. Both approaches might require numerical evaluation. One question that computation of conditional and marginal moments could be used to address is whether there exist

parameter values in the two models that produce some type of "equivalence" in the values computed. For example, although the systematic components of conditional models would differ between the two model structures (else there would have been no motivation for developing parameterized families of link functions in the first place), one might ask the question of whether the systematic components of marginal models could be made the same or quite similar. Or, one could average over covariate values to determine whether the overall marginal means of the two models could be matched and, if so, what the effect on marginal variances is.

Simulation of data sets under the two models would be facilitated by first having some type of "matching" of at least marginal moments, possibly averaged over covariate values. If this is possible, data sets could be simulated under the two models using parameter values that produce such matching, and those data sets examined to determine whether the data structures produced could be distinguished in a systematic manner. For example, if the models could be arranged so that marginal expectations averaged over covariate values are the same, one could examine empirical distributions of conditional expectations averaged over covariate values.

7. Having support of density or mass functions that depends on parameter values is usually taken to be a "red flag" for the application of asymptotic theory, since a portion of one of the basic regularity conditions is that the observable random variables have common support that does not depend on the parameter. The situation here is a bit more involved, however. For the regression model itself, based on observable random variables having binomial probability mass functions, there is no reason $\lambda < 0$ should cause any problems with asymptotic results. Thus, for forming intervals for regression parameters β_0 and β_1 and even the link function parameter λ there would be no increase in

concern with using typical Wald theory caused by the restriction on support of the tolerance distribution. That is, asymptotic normality of maximum likelihood estimates of parameters in the model is not affected by this restriction. In terms of making inference about the tolerance distribution, however, this problem becomes important. For example, even if the point estimate of λ is positive, an interval estimate might include negative values, implying that the tolerance distributions "included" in such an interval do not all have the same support. This would certainly not be a pleasing occurrence. Interval estimation of quantiles of the tolerance distribution using the delta method to transform the inverse information matrix into standard errors for an estimated quantile could be completely destroyed; consider, for example, and interval for the 0.95 quantile of the tolerance distribution in which a portion of that interval lies outside of the support dictated by the maximum likelihood estimate $\hat{\lambda}$.