

**THE UNIVERSITY OF MANITOBA**

***A Retrospective Analysis of Survival  
Data of Snapping Turtle Embryos using  
Generalized Linear Models***

**BY**

**AMELIA TERESA SHEOCHARAN**

**A PRACTICUM SUBMITTED TO THE FACULTY OF GRADUATE  
STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF SCIENCE**

**DEPARTMENT OF STATISTICS**

**June 2005**



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

0-494-08955-5

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*ISBN:*

*Our file* *Notre référence*

*ISBN:*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

**THE UNIVERSITY OF MANITOBA**  
**FACULTY OF GRADUATE STUDIES**  
\*\*\*\*\*  
**COPYRIGHT PERMISSION**

**“A Retrospective Analysis of Survival Data of Snapping Turtle Embryos using Generalized Linear Models”**

**BY**

Amelia Teresa Sheocharan

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of  
Manitoba in partial fulfillment of the requirement of the degree  
Of  
MASTER OF SCIENCE**

Amelia Teresa Sheocharan © 2005

**Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.**

**This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.**

## ***Abstract***

The effect of temperature, moisture and site on the proportion of turtle eggs that survived after hatching was analyzed. Preliminary main effect plots showed that temperature and moisture had no effect on the survival of a turtle egg. On the other hand site showed some differences. Next, the interaction effects were then examined using interaction plots. Here it was found that site by temperature showed some slight differences.

The data were analyzed using a newly developed technique called Generalized Linear Models. After the analysis was completed it was found that indeed site had an effect on the survival rate of turtle eggs.

## ***Acknowledgements***

First and foremost I would like to thank my supervisor Dr. K. Mount for taking me on as one of his students, and for all his help, support and patience. I would also like to thank my committee, Dr. B. Macpherson and Mr. W. Falk for taking time out of their busy schedule to proof read the practicum and offer their suggestions. I would also like to thank the Department of Statistics, Dr. S. Cheng for giving me the many opportunities while completing my project.

Finally, I would like to thank all my friends and family for their continued love and support.

# ***Table of Content***

<b>Introduction.....</b>	<b>1</b>
<b>Chapter 1: Overview.....</b>	<b>2</b>
1.1 The Objective.....	2
1.2 The Experiment.....	2
1.3 Design of the experiment.....	5
1.4 Initial plots for the main effects.....	6
1.5 Preliminary interaction plots.....	8
<b>Chapter 2: Analysis of the Temperatures.....</b>	<b>14</b>
2.1 Introduction.....	14
2.2 Iterative method for parameter estimation in Generalized linear models.....	17
2.3 Fisher scoring method.....	19
2.4 Quasi – likelihood .....	20
2.5 Generalized linear model inference.....	20
2.6 Residual.....	22
2.7 Temperature Analysis.....	23
<b>Chapter 3: Analysis of the Moistures.....</b>	<b>29</b>
3.1 Introduction.....	29
3.2 Moisture Analysis.....	31
3.2.1 Penalized Quasi likelihood.....	36
3.2.2 Fitting the Model.....	36

<b>Chapter 4: Analysis of the Sites.....</b>	<b>41</b>
4.1 Introduction.....	41
4.2 Sites Analysis.....	42
4.2.1 Fixed Sites.....	43
4.2.2 Random Sites.....	54
4.3 Best linear unbiased predictors.....	62
4.3.1 Broad Inference.....	64
4.3.2 Narrow Inference.....	65
<b>Chapter 5: Summary.....</b>	<b>69</b>
5.1 Summary.....	69
<b>Appendix A.....</b>	<b>71</b>
Chapter 1: Proportion Survived and Other Descriptive Statistics.....	71
Chapter 4: Proportion of turtles eggs survived by clutch.....	72
<b>Appendix B: SAS Programs.....</b>	<b>73</b>
<b>Appendix C: Original Case Study.....</b>	<b>85</b>
<b>Bibliography.....</b>	<b>93</b>

## ***List of Figures***

Figure 1.1 Map of Ontario.....	3
Figure 1.2 Plot of Proportion Alive by Temperature.....	6
Figure 1.3 Plot of Proportion Alive by Moisture.....	7
Figure 1.4 Plot of Proportion Alive by Site.....	8
Figure 1.5 Plot of Proportion Alive by Temperature and Moisture.....	9
Figure 1.6 Plot of Proportion Alive by Site and Moisture.....	10
Figure 1.7 Plot of Proportion Alive by Temperature and Site.....	11
Figure 1.8 Plot of Proportion Alive by Clutch .....	12
Figure 1.9 Plot of Proportion Alive by Clutch and Temperature.....	13
Figure 2.1 Front view of an incubator at the whole plot level.....	24
Figure 2.2 Residual Plot of Y by Predicted Y.....	28
Figure 3.1 Front view of an incubator at the sub plot level.....	32
Figure 3.2 Residual Plot of Y by Predicted Y.....	40
Figure 4.1 Top view of a tray in an incubator .....	42
Figure 4.2 Residual Plot of Y by Predicted Y.....	53
Figure 4.3 Residual Plot of Y by Predicted Y.....	62

## ***List of Tables***

Table 1.1 Number of eggs collected per site.....	4
Table 2.1 Table of functions for the three distributions.....	16

## ***Introduction***

In 1988 Michele Bobyn, a student at the University of Guelph, conducted an experiment to help her determine the incubation conditions needed to influence embryonic survival in snapping turtles. The research in this practicum re-evaluates the same study but uses modern methods to analyze the 1988 data.

The objective of this practicum is to determine the effect of temperature, moisture and site on the embryonic survival of a snapping turtle egg. The analysis used is a newly developed method called generalized linear model.

A generalized linear mixed model will be fit and used to help determine what conditions are needed to have the highest survival rate of a turtle egg and estimate the probability reflected in those conditions.

In chapter 1 we focus on the factors our researcher is considering and make some initial main effect and interaction evaluations.

Chapter 2 examines the temperature effects and introduces new concepts needed to evaluate the data.

Chapter 3 builds on the temperature model and examines the added effects in the model when the factor moisture is introduced.

Chapter 4 introduces the last effect site and uses all our data to fit a model that best reflects the data. The model is then used to make predictions for the turtles using best linear unbiased techniques.

Finally in chapter 5 we summarize our conclusions and give final remarks.

# ***Chapter 1: Overview***

## **1.1 THE OBJECTIVE**

The following information was provided through a case study report written by the experimenter.

The experiment was conducted in June 1988 by Michele Bobyn at the University of Guelph under the supervision of Dr. R.J. Brooks.

The objective was to:

'determine the effect of temperature, moisture and site on the embryonic survival of a snapping turtle egg and to determine how embryonic survival varies from clutch to clutch.'[3]

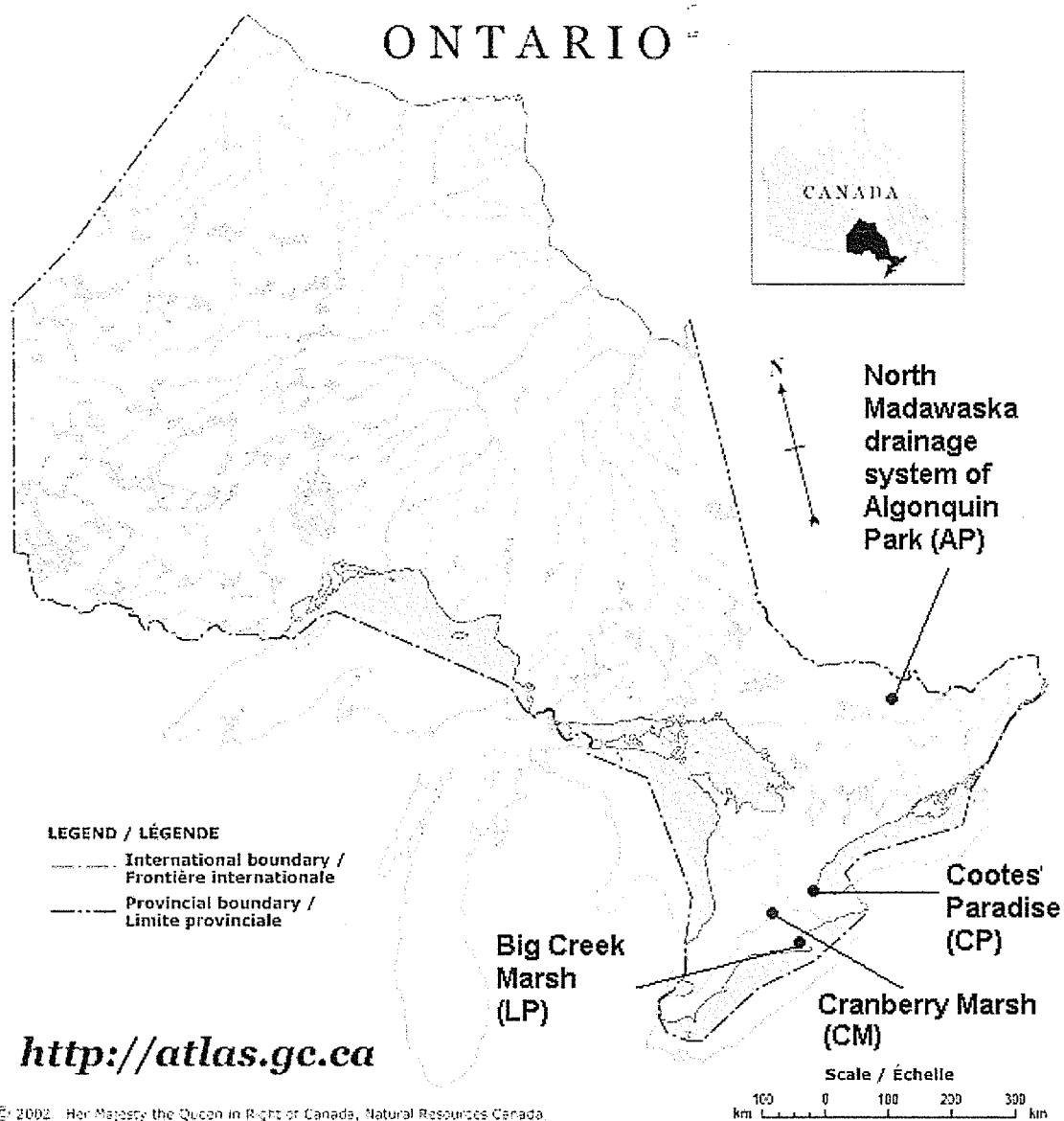
## **1.2 THE EXPERIMENT**

Snapping turtle eggs were collected from four Ontario nesting sites:

- I. North Madawaska drainage system of Algonquin Park (AP)
- II. Cootes' Paradise near Hamilton (CP)
- III. Big Creek Marsh near Long Point (LP)
- IV. Cranberry Marsh near Ajax (CM).

The sites are shown on figure 1.1.

Figure 1.1:



© 2002 Her Majesty the Queen in Right of Canada, Natural Resources Canada  
 Sa Majesté la Reine du chef du Canada, Ressources naturelles Canada

Each egg was labelled with:

- its location (AP, CP, LP, CM)
- clutch identification
- egg number (#1 = last laid egg).

For transport the eggs were arranged in a single layer plastic shoe box and covered with a mixture of vermiculite (*a type of mineral*) and water. The shoe box was covered with aluminium foil to prevent dehydration. All eggs were maintained at a constant 20°C before being placed into an incubator.

For the experiment there were six incubators used, two set at each of three temperatures: 21°C, 25°C and 29.5°C. A total of 720 eggs were collected. The following table shows how many eggs were collected from each site.

Table 1.1:

<b>Site</b>	<b>No. of clutches</b>	<b>No. of eggs</b>
AP	6	212
CP	5	183
LP	5	154
CM	5	171
<b>TOTAL</b>	21	720

The eggs were then randomly assigned to a tray. Each tray had a total of 40 eggs with a maximum of 2 eggs from each clutch at each site.

The trays were then randomly assigned to one of two moistures: dry or wet.

A total of three trays were placed in each incubator. The combinations of moistures were either two wet and one dry or two dry and one wet tray per incubator.

When the first sign of hatching occurred in an incubator all the eggs were removed at that temperature and placed in a glass jar.

Less than a week into the start of the incubation period one of the incubators at 29.5°C malfunctioned and 120 eggs were lost. This left five incubators for the study.

The response variable, survival, was measured two months after the egg hatched. Survival was scored a 1 or 0 depending on whether the turtle lived or died.

### **1.3 DESIGN OF THE EXPERIMENT**

The design of the experiment followed a split- split plot layout. At the whole plot level the factor is temperature and the experimental units are the incubators. Two incubators were randomly assigned to each temperature level.

At the split plot level the factor is moisture and the experimental units are the trays within incubators. At this point the trays were assigned to one of two moisture levels; wet or dry.

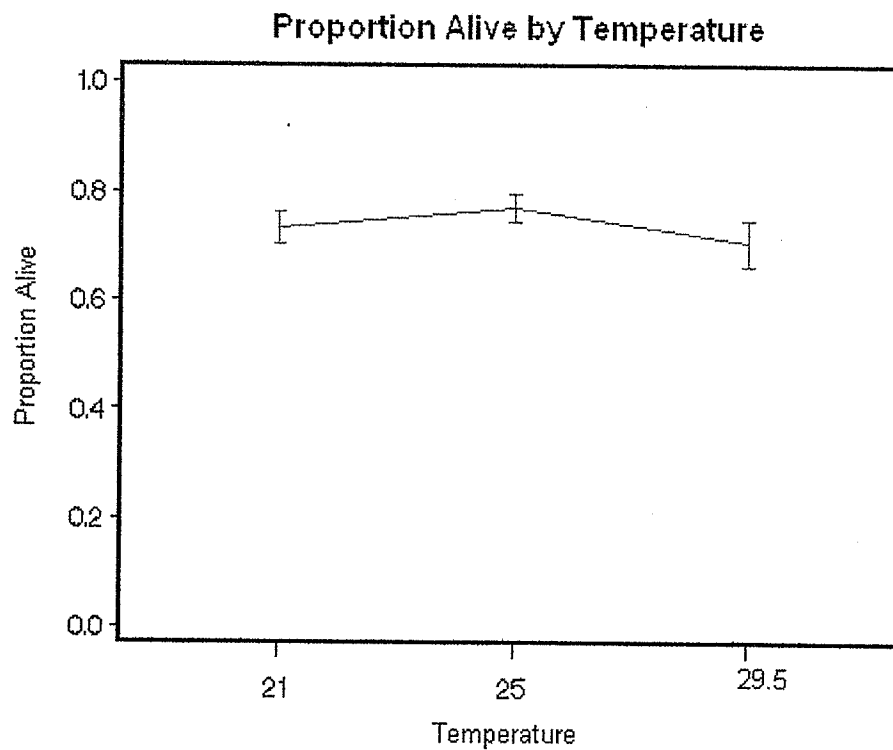
At the split- split plot level the factor is site and the experimental units are the eggs. For the experiment approximately 600 eggs were used.

## 1.4 INITIAL PLOTS FOR THE MAIN EFFECTS

Before any formal data analysis is done, we are going to look at main effect and interaction plots with plus and minus one standard error bands for each data point.

The first factor, temperature, has 3 levels: 21°C, 25°C, and 29.5°C.

Figure 1.2:

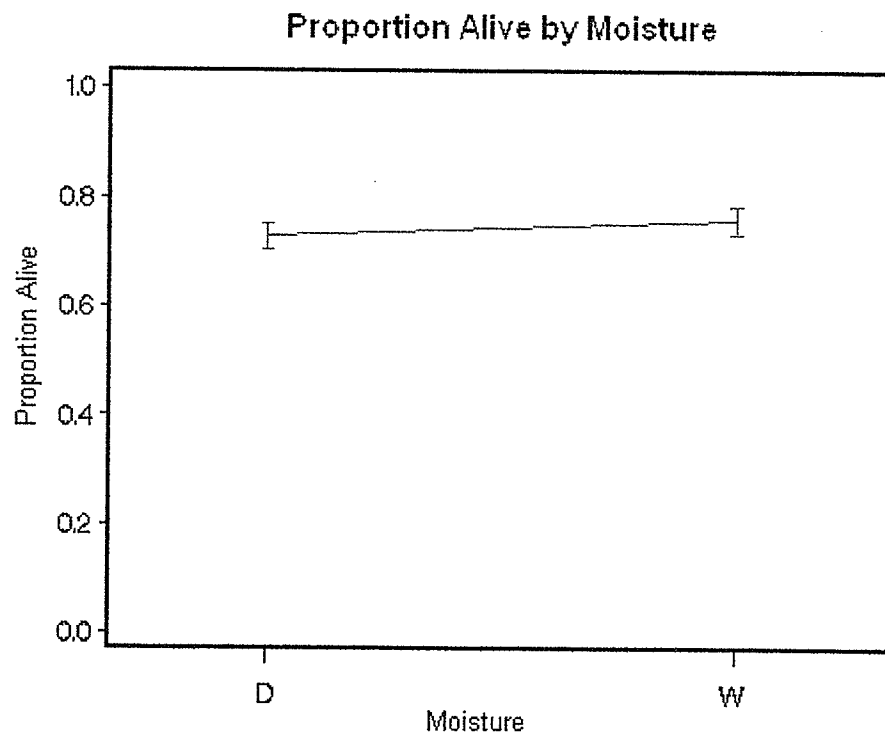


As we can see there is very little change in the proportion of turtles that survived as the temperature increased from 21°C to 29.5°C. This tells us that the main effect of temperature has very little effect on the survival of a turtle egg. We notice that at temperature 29.5°C there is a larger standard error; that is because at 29.5°C one of the incubators malfunctioned and 120 eggs were lost. At each

21<sup>0</sup>C and 25<sup>0</sup>C we are looking at the number of eggs that survived out of 240, whereas at 29.5<sup>0</sup>C we are looking at the number of eggs that survived out of 120.

Next, we looked at the factor moisture.

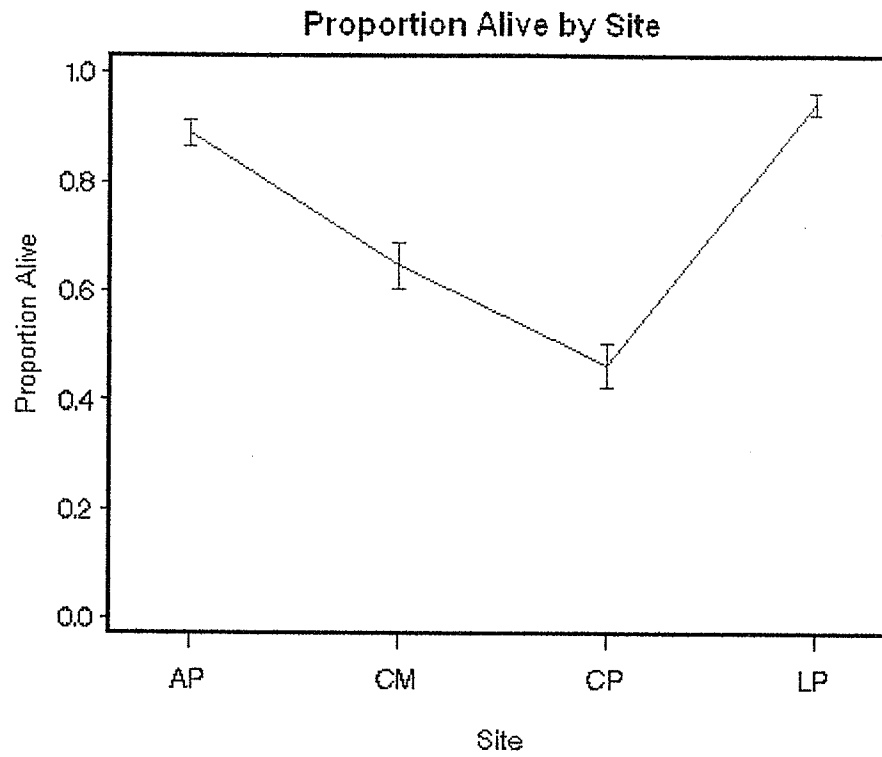
Figure 1.3:



As we can see moisture shows very little difference between wet and dry, which tells us that the factor moisture has no notable effect on the survival of a turtle.

The next factor is site.

Figure 1.4:



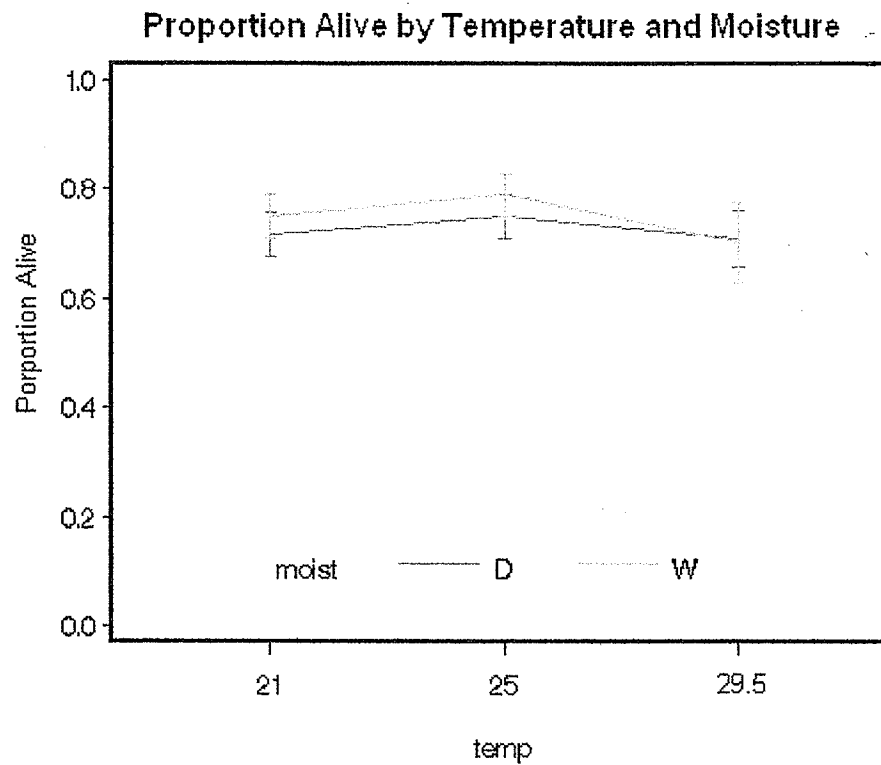
We notice there is a considerable difference as to where the turtle egg came from and whether it survived. It seems Big Creek Marsh (LP) had the highest mean survival rate and Cootes' Paradise (CP) the lowest.

## 1.5 PRELIMINARY INTERACTION PLOTS

Next, we consider interactions between factors.

First we look at temperature by moisture.

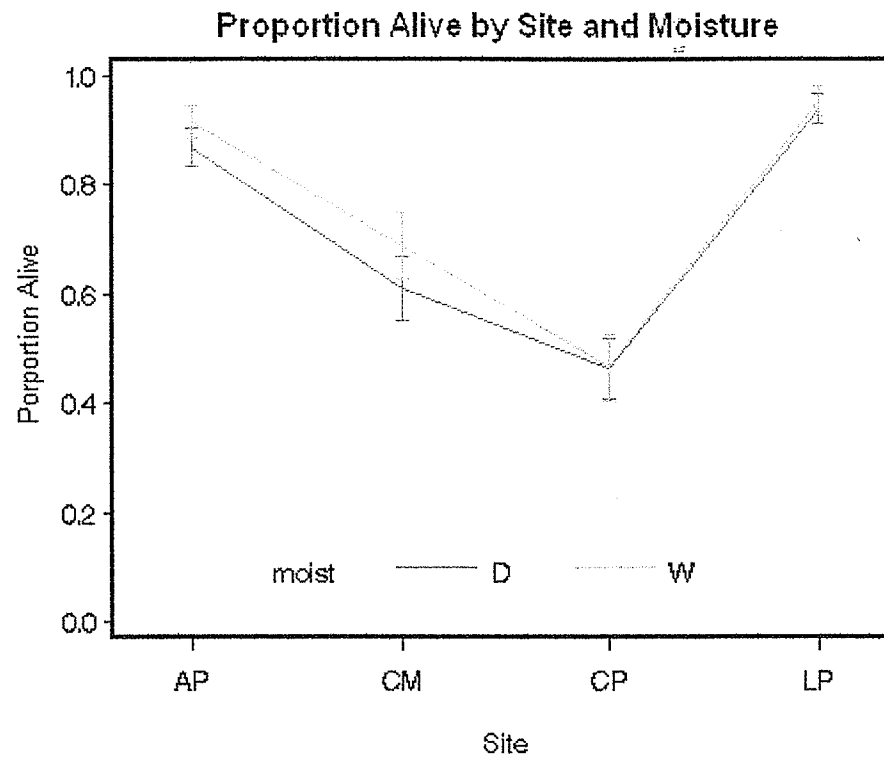
Figure 1.5:



As we can see the interaction plot of temperature by moisture shows no interaction effect.

Next, we look at site by moisture.

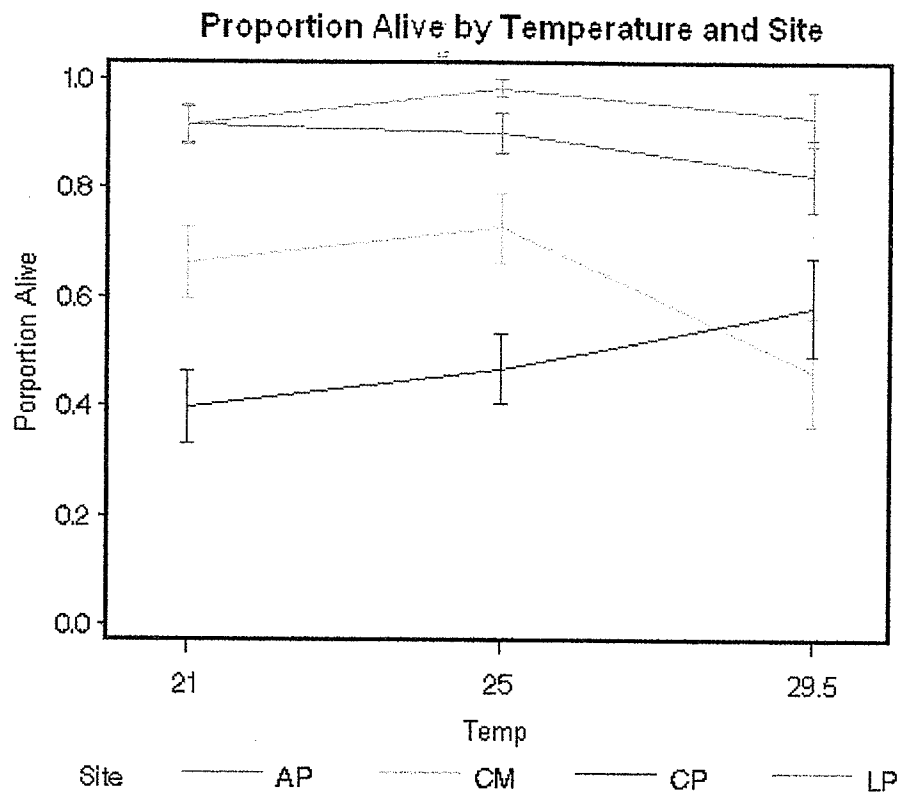
Figure 1.6:



Again we see no interaction between these two factors.

Lastly, we look at the temperature and site interaction to see if there are any effects.

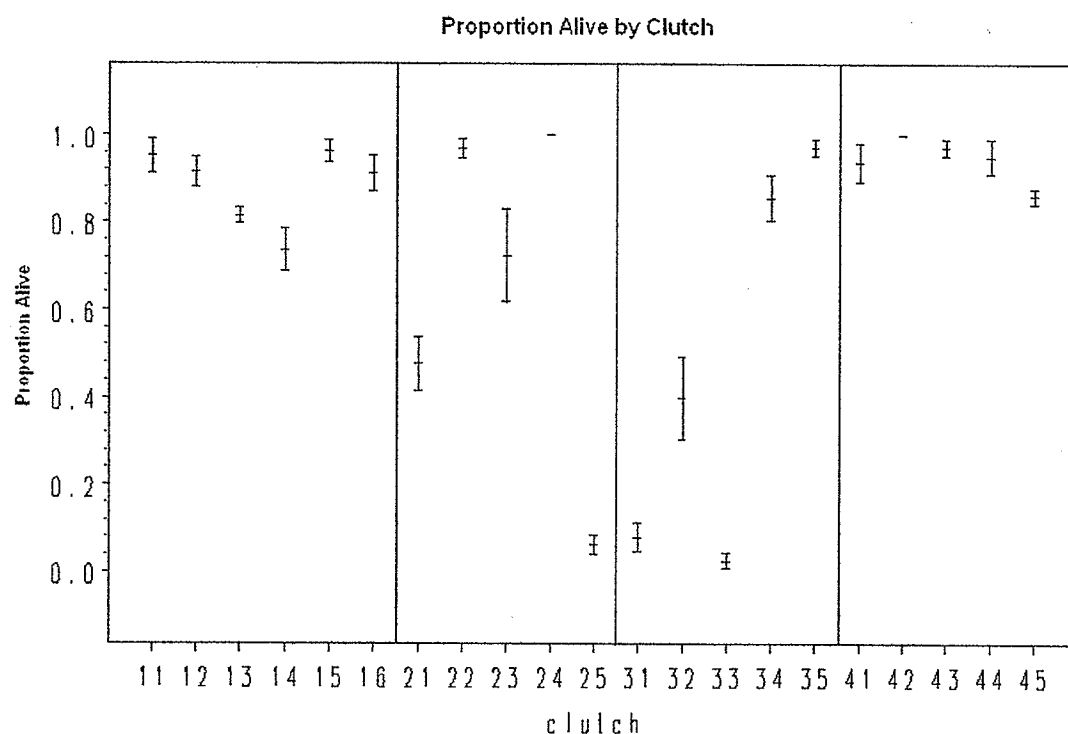
Figure 1.7:



As we can see there appear to be some differences here. As we continue we will revisit this relationship.

At this time we will look at the clutch to clutch variation. To do this we have plotted the proportion of turtle eggs that survived in each clutch from our four sites.

Figure 1.8:



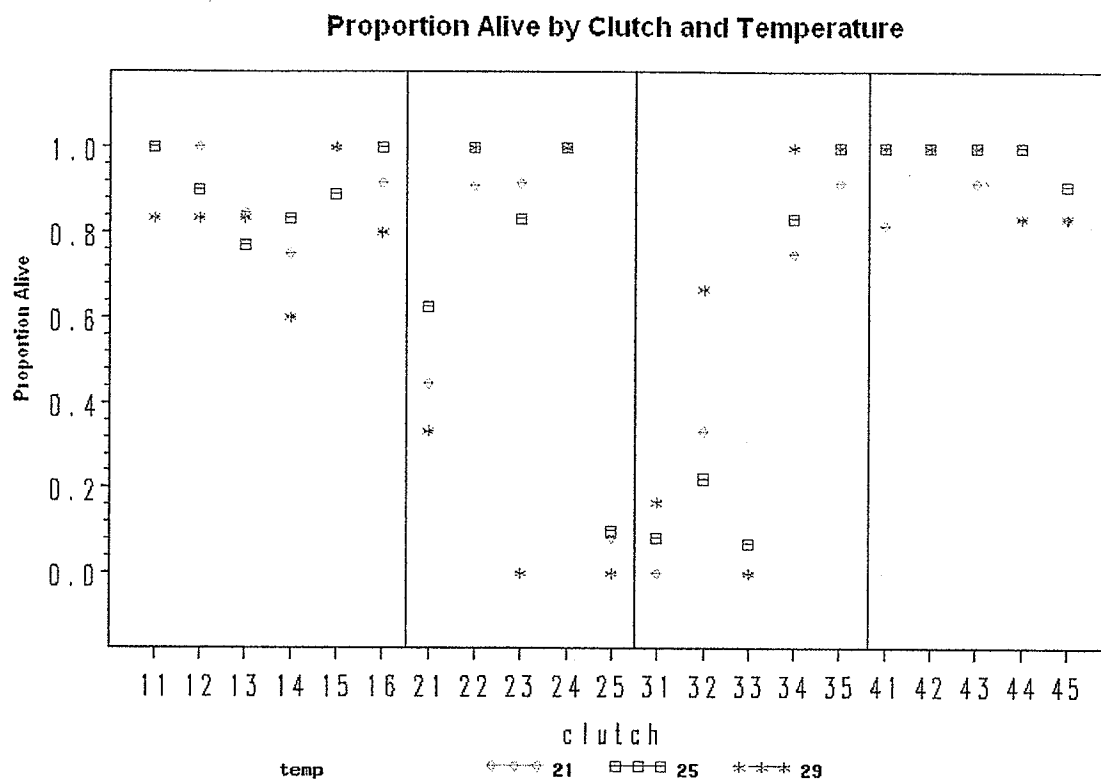
**LEGEND:**

Clutch #11 – 16: Site: AP  
 Clutch #21 – 25: Site: CM  
 Clutch #31 – 35: Site: CP  
 Clutch #41 – 45: Site: LP

From this we see there is similar variation between sites AP and LP as well as sites CM and CP. Overall, we see that sites CM and CP are more variable than sites AP and LP.

The next graph looks at the proportion of turtle eggs that survived from each clutch at each temperature level.

Figure 1.9:



**LEGEND:**

Clutch #11 – 16: Site: AP  
 Clutch #21 – 25: Site: CM  
 Clutch #31 – 35: Site: CP  
 Clutch #41 – 45: Site: LP

From this graph we do not see any patterns between clutch and temperature.

## ***Chapter 2: Analysis of the Temperatures***

### **2.1 INTRODUCTION**

At the first level, the whole plot, one looks at the effect temperature has on the survival of a turtle egg.

If the response variable were continuous we might assume that it followed a normal distribution therefore a linear model or general linear model could be used. Because our data are binomially distributed a generalized linear model is appropriate at this stage.

A linear model has model equation  $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ , where the  $E(\underline{Y}) = X\underline{\beta}$ . It postulates a linear relationship between a dependent or response variable  $\underline{Y}$ , and a linear combination of fixed predictor variables. The  $\varepsilon$  stands for the error i.e. variability in  $\underline{Y}$  that cannot be accounted for by the predictors. The expected value of an error is assumed to be zero and it is independently normally distributed with constant variance i.e.  $\underline{\varepsilon} \sim N(0, \sigma^2 I)$ .

To estimate the unknown parameters we use the method of least squares. Because we assume the error terms are normally distributed we can carry out tests on the parameters. In addition confidence intervals for the parameters, and confidence intervals for the mean of the response variable are obtainable.

In the general linear model,  $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ ,  $E(\underline{Y}) = X\underline{\beta}$  as above, and  $\underline{\varepsilon}$  has mean zero but a more general variance covariance structure ie.  $\Sigma$ . As before, the dependent variables are expected to follow the normal distribution. To estimate the unknown parameters we now use the method of general least squares.

A generalized linear model, with form  $g(\mu) = X\beta$  is a generalization of the general linear model. It was developed for data that do not follow the assumptions of a general linear model. For example, we have independent response variables  $y_1, \dots, y_n$  with means  $\mu_1, \dots, \mu_n$ . The response variable may or may not be a continuous variable; instead it could be a count. A generalized linear model can be used in two situations:

- i. for dependent variables which are discrete random variables
- ii. for dependent variables, which are not linearly related to the predictor variables i.e. data that needs to be transformed so that a function of the mean of the response variable is linearly related to the predictor variables.

The generalized linear model has three aspects:

First, it extends linear models to the situation where response variables are members of the exponential family. The exponential family includes normal, binomial, Poisson, Geometric, negative binomial, exponential, gamma and inverse normal distributions. Members of the exponential family of distributions have probability density function for a response  $y$  that can be expressed in the form:

$$f_y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}$$

where:

$\theta$  is the natural location parameter. It is a function of the mean  $\theta(\mu)$ , where  $\mu$  is the mean of the response.

$\phi$  is a scale or dispersion parameter.

$a(\cdot), b(\cdot)$  and  $c(\cdot)$  are specific functions.

The variance,  $v(\mu)$ , is a function of  $\mu$  rather than  $\theta$ .

Table 2.1, taken from SAS System for Mixed Models, is a table of functions for some common members of the exponential family.

Table 2.1:

**Table of functions for the three distributions [5]**

	Poisson	Normal	Bernoulli
Mean(y)	$\lambda$	$\mu$	$\pi$
Var(y)	$\lambda$	$\sigma^2$	$[\pi(1-\pi)]$
$\theta(\mu)$	$\log_e(\lambda)$	$\mu$	$\log_e[\pi/(1-\pi)]$
$a(\phi)$	1	$\sigma^2$	1
$v(\mu)$	$\lambda$	1	$\pi(1-\pi)$

Secondly, as in the two previous models, it has a set of parameters  $\underline{\beta}$  and explanatory variables  $x' = [x_1, \dots, x_p]$

Thirdly, there is a link function  $g$  such that:  $g(\mu_i) = X_i \underline{\beta}$  where  $\eta_i = g(\mu_i)$

The link function is a function of the mean  $\mu$ . It connects the mean of the raw data to the natural parameters to give us the basic form of the generalized linear model:  $\eta = X\beta$ . Various link functions can be chosen, depending on the

assumed distribution of the  $y$  variable e.g. logit link:  $\log\left(\frac{\pi}{1-\pi}\right)$ .

In the analysis of this chapter we will use a logit link function since the data we are dealing with follow a binomial distribution.

In generalized linear models, iterative methods are needed to solve for the parameter estimates.

## 2.2 ITERATIVE METHOD FOR PARAMETER ESTIMATION IN GENERALIZED LINEAR MODELS.

The next two sections are a summary of the iterative method for parameter estimation from Generalized Linear and Mixed Models by Searle and McCulloch [8].

The method of maximum likelihood is the basis for parameter estimation in generalized linear models. However, the actual operation of maximum likelihood results in an algorithm based on iteratively weighted least squares.

The likelihood function is defined as the joint density function of  $n$  random variables  $f(\underline{x}, \theta)$ . It is considered a function of  $\theta$  and can be denoted as  $L(\theta, \underline{x})$ .

Given  $\underline{x}$  we want to maximize  $L(\theta, \underline{x})$  or in this case  $L(\underline{y}, \underline{\beta})$ . We set the derivative of the log-likelihood  $L(\underline{y}, \underline{\beta})$  to 0 and solve for  $\hat{\beta}$ .

In matrix notation the maximum likelihood equations (also called the score equations) can be written as

$$X'W\Delta\underline{y} = X'W\Delta\underline{\mu}$$

where:

$W = \{w_i\}$  is a nxn diagonal matrix with elements  $w_i = [v(\mu_i)g_\mu^2(\mu_i)]^{-1}$  and

$g_\mu(\mu_i) = \frac{\partial \eta_i}{\partial \mu_i}$ .  $v(\mu_i)$  is the variance function for the  $i$ th observation. We note

that  $\mu_i$  is  $\pi_i$ . The weights ( $w_i$ ) depend on the parameters  $\underline{\mu_i}$ . We can re-write  $w_i$

as  $\frac{1}{\text{var}(y_i)} \left( \frac{\partial \mu_i}{\partial \eta} \right)^2$ . For the turtle data, the variable  $y$  follows a binomial

distribution. Thus  $\text{var}(y_i)$  equals  $[\pi_i(1-\pi_i)]$ . Therefore the  $i$ th diagonal element

of the  $W$  matrix is  $w_i = \frac{1}{\pi_i(1-\pi_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$ .  $\Delta = \{g_\mu(\mu_i)\}$  is an nxn diagonal matrix.

From this we see that  $W$  and  $\Delta$  both depend on the mean  $\underline{\mu_i}$ . Therefore  $W$ ,  $\Delta$  and  $\underline{\mu}$  (or  $\underline{\pi}$ ) all involve the unknown parameter  $\underline{\beta}$  through the link function. Because the maximum likelihood equation is a nonlinear function of  $\underline{\beta}$  we cannot solve this equation analytically. Therefore we use the Fisher scoring method.

## 2.3 FISHER SCORING METHOD

Solutions for the maximum likelihood equation for  $\underline{\beta}$  are performed by an iterative weighted least squares method. This can be derived as an example of the use of Fisher scoring. "Fisher scoring, the method used by SAS, is an iterative method for maximizing a likelihood and it takes on the form:

$$\theta^{m+1} = \theta^{(m)} + I(\theta^{(m)})^{-1} \frac{\partial l}{\partial \theta} \Big|_{\theta=\theta^{(m)}}$$

where (m) represents the mth iteration,  $I(\theta)$  is the information matrix and  $\theta$  is the entire parameter vector" [8].

If we rewrite the previous equation in the context of our situation we have:

$$\underline{\beta}^{(m+1)} = \underline{\beta}^{(m)} + \left( X'W^{(m)}X \right)^{-1} X'W^{(m)}\Delta^{(m)} \left( \underline{y} - \underline{\mu}^{(m)} \right)$$

For this iterative procedure we need starting value estimates for the parameter  $\underline{\beta}^{(0)}$ . Once we have this value we can obtain starting values for  $\mu, W, \Delta$  (denoted by  $\mu^{(0)}, W^{(0)}, \Delta^{(0)}$ ). These are all part of our iterative equation.

We then use these estimates to revise our equation to solve for the next parameter estimate. We linearize the model about these new values and linear least squares is applied again to find a second set of estimates. This procedure is repeated until the desired degree of convergence is obtained.

## 2.4 QUASI – LIKELIHOOD

The following is adapted from Generalized Linear Models by Myers, Montgomery and Vining [9].

Sometimes we are dealing with situations in which a modified version of maximum likelihood is needed. One such situation is when the responses are correlated. Wedderburn (1974) developed the concept of quasi - likelihood for such situations. The main point is that in using maximum likelihood above (i.e. sec. 2.2 & 2.3) we are only using the first two moments of the distribution of the response. The score equations are replaced by the equations  $F'V^{-1}(\underline{y} - \underline{\mu}) = \underline{0}$

where F is a matrix of derivatives  $\frac{d\mu}{d\beta}$ . If the responses are independent this

equation can eventually be written in the form  $X'\Delta V^{-1}(\underline{y} - \underline{\mu}) = \underline{0}$  where

$$\Delta = \text{diag} \left[ \frac{\partial \mu_i}{\partial \eta_i} \right].$$

## 2.5 GENERALIZED LINEAR MODEL INFERENCE

Inference for a generalized linear model starts with its deviance.

The deviance (D) is defined as:

$$D(\beta) = -2 \ln \left[ \frac{L(\underline{\beta})}{L(\underline{\mu})} \right]$$

where:  $L(\underline{\beta})$  is the likelihood for the fitted logistic model, the reduced model, and

$L(\underline{\mu})$  is the likelihood for the saturated model. The saturated model is regarded

as providing a complete description of the data. A probability is given for each unit at this level. The reduced model reparameterizes the saturated model such that we have fewer parameters than the saturated model. The deviance allows us to compare the reduced with the saturated model to see if the reduced model is an adequate fit.

If the reduced model fits the data then the deviance is approximately distributed as a  $\chi^2$  with  $n-p$  degrees of freedom, where  $n$  is the number of observations and  $p$  is the number of distinct parameters in  $\beta$ .

The deviance can be used for two purposes:

1. To assess goodness of fit of the reduced model, by comparing the calculated deviance to the critical value of the appropriate  $\chi^2$  distribution.
2. Hypothesis testing, which consists in thinking of each hypothesis in terms of a model and comparing the goodness of fit statistic for the full model and the reduced model. For example,  $\beta$  is broken down into two sub vectors  $\beta_1$  and  $\beta_2$  so we can write  $X\beta$  as  $X\beta = X_1\beta_1 + X_2\beta_2$ . Now, we use the difference between the deviance of the full model and the deviance of the reduced model (i.e. fitting  $X_1\beta_1$  alone) to test  $H_0 : \beta_2 = 0$ .

The likelihood ratio statistic is approximately distributed as  $\chi^2$  with  $p-p_1$  degrees of freedom, where  $p_1$  is the number of distinct parameters in  $\beta_1$ .

Hypothesis testing can also be done using Wald, or score tests.

In addition to testing hypotheses and assessing goodness of fit we are also able to calculate confidence intervals for the parameters, and confidence

intervals for the expected value of the response variable in a generalized linear model.

## 2.6 RESIDUALS

After we fit the generalized linear model we should check to see if there are any problems with the model.

Similarly to linear regression, we will look at residual plots.

For fixed effects there are three types of residuals available in SAS for model checking; they are Pearson, deviance and raw residuals.

The Pearson residuals are defined as  $r_{p_i} = \frac{(y_i - \hat{\mu}_i)\sqrt{w_i}}{\sqrt{V(\hat{\mu}_i)}}$  where  $w_i$  is the

weight matrix. The  $i$ th residual is the  $i$ th contribution to the Pearson Chi-square.

The deviance residual is defined as  $r_{d_i} = (\text{sign}(y_i - \hat{\mu}_i))\sqrt{d_i}$  where  $d_i$  is the square root of the  $i$ th contribution to the total deviance. In other words

$$\sum d_i = D(\beta).$$

Lastly, the raw residuals  $r_i$  are given by,  $r_i = y_i - \hat{\mu}_i$ , where  $y_i$  is the response and  $\hat{\mu}_i$  is the corresponding predicted mean.

According to Myers, Montgomery & Vining [9] the raw residuals are not appropriate for generalized linear models because the  $\text{var}(y_i)$  is not constant.

McCullagh & Nelder [7] recommend plotting the deviance residuals.

The residual plot is a scatterplot of standardized deviance residuals against the estimated linear predictor  $\hat{\eta}$  or against the fitted values.

The fitted values are transformed to a constant scale based on the error distribution. For the turtle data the distribution was assumed to be binomial.

Binomial errors are transformed using  $2\sin^{-1}\sqrt{\hat{\mu}}$ .

After constructing a residual plot you want to look for patterns. Examples of patterns seen in the residual plot that indicate a problem with the model include curvature in the residual plot, and a systematic change in the range of the residuals with the fitted values.

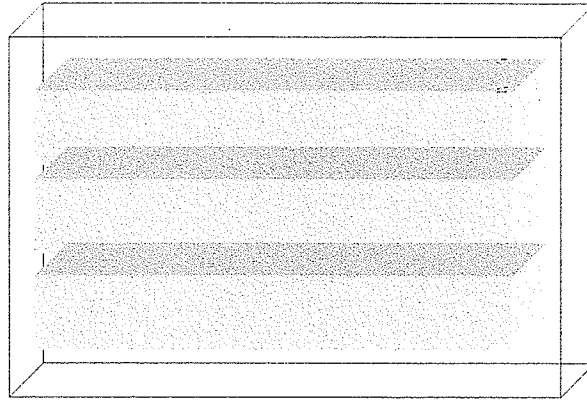
Curvature in the residual plots may be the result of choosing the wrong link function, wrong choice of scale of one or more covariates, or omission of a quadratic term in a covariate. Residual plots are also used to help detect outliers or influential observations.

## 2.7 TEMPERATURE ANALYSIS

Utilizing the turtle data at the whole plot level, we consider the effect of temperature (denoted as  $\tau_i$ ). Temperature has three levels (21°C, 25°C and 29.5°C, so  $i = 1, 2, 3$ ). At each temperature we have two incubators, except at temperature 29.5°C where one incubator malfunctioned and all the eggs were lost.

Figure 2.1 is a front view of an incubator. The incubator is a visual representation of the experimental unit at this stage of the analysis.

Figure 2.1:



The response variable  $Y$  is the survival of a turtle egg. It is defined as:

$$Y = \begin{cases} 1, & \text{if egg survives (success)} \\ 0, & \text{if egg dies (failure)} \end{cases}$$

We assume that  $Y$  follows a Bernoulli distribution. Since this is a member of the exponential family, a generalized linear model can be used to analyze the data at this stage.

For the turtle data let  $Y_{ij}$  = number of turtle eggs that survived in incubator  $j$  at temperature  $i$ . Then

$$Y_{ij} \sim \text{Binomial}(n_{ij}, \pi_i)$$

where:

$\pi_i$  is the probability that a turtle egg at the  $i$ th temperature successfully hatches and  $n_{ij}$  denotes the number of turtle eggs at temperature  $i$  in incubator  $j$ .

A generalized linear model at the whole plot level for the turtle data can be defined as follows:

For binomially distributed data a logit link function is used. Hence, we write the model as  $\log\left(\frac{\pi}{1-\pi}\right) = X\beta$ . Here we have two incubators ( $j = 1, 2$ ) randomly assigned to the three temperature levels ( $i = 1, 2, 3$ ) and so

$$\underline{\pi} = \begin{pmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \\ \pi_{22} \\ \pi_{31} \end{pmatrix}, \text{ and } \log\left(\frac{\pi}{1-\pi}\right) \text{ is the vector of size } 5 \times 1 \text{ representing the mean}$$

logits of turtle eggs successfully hatching in  $j$ th incubator at temperature  $i$ . We recall that one incubator failed at temperature  $29.5^{\circ}\text{C}$  and all the eggs were lost.

The  $X$  and  $\beta$  matrices are defined as follows;

$$\beta = \begin{pmatrix} \mu^* \\ \tau_1^* \\ \tau_2^* \end{pmatrix}$$

To assess the temperature effect SAS uses a reference cell formulation ie. all means are written in terms of their differences from a certain cell mean which is referred to as the reference cell. In the case of the temperature effect the reference cell is identified as temperature  $29.5^{\circ}\text{C}$ . Using the reference cell formulation the parameters at the whole plot level can be defined as follows:

$\mu^*$  is the mean logit  $\left(\log\left(\frac{\pi}{1-\pi}\right)\right)$  at  $29.5^{\circ}\text{C}$ .

$\tau_1^*$  is the mean logit at  $21^{\circ}\text{C}$  – mean logit at  $29.5^{\circ}\text{C}$

$\tau_2^*$  is the mean logit at  $25^{\circ}\text{C}$  – mean logit at  $29.5^{\circ}\text{C}$

The design matrix is:  $X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$  : Thus the expected value form of the

model equation is  $\log\left(\frac{\pi}{1-\pi}\right) = X\beta$ .

Earlier, when we looked at our main effect plots we saw that temperature had no effect on the survival of a turtle egg. To look at this more formally we use PROC GENMOD in SAS.

PROC GENMOD is designed to fit generalized linear models using maximum likelihood theory and iterative methods to solve for the parameter estimates.

The following selected output was obtained using SAS.

First, we will assess the goodness of fit of our model. The deviance given to us in SAS will be used to determine the fit.

#### Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	0.5509	0.2754

The  $P(\chi^2_2 > 0.5509) = 0.7592$ , suggests that the model is a good fit.

Next we test the hypotheses:

$$H_{0_i} : \tau_i^* = 0$$

vs

$$H_{a_i} : \tau_i^* \neq 0$$

where  $\tau_i^*$  ( $i=1,2$ ) is previously defined.

### Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	0.8473	0.1992	18.09	<.0001
temp 21	1	0.1431	0.2465	0.34	0.5616
temp 25	1	0.3191	0.2504	1.62	0.2024
temp 29.5	0	0.0000	0.0000		
Scale	0	1.0000	0.0000	1.00	1.0000

Using the Chi-squared test, from SAS we get  $\chi_1^2 = 0.34$ . The corresponding p-value 0.5616 is not significant which suggests that temperature 21°C has the same impact as temperature 29.5°C on the probability of survival of a turtle egg. A similar result was obtained for 25°C.

Alternatively we can use the likelihood ratio test from SAS to again look at the hypothesis that temperature has no effect on the survival of a turtle egg versus the alternative that temperature does have an effect on the survival of a turtle egg.

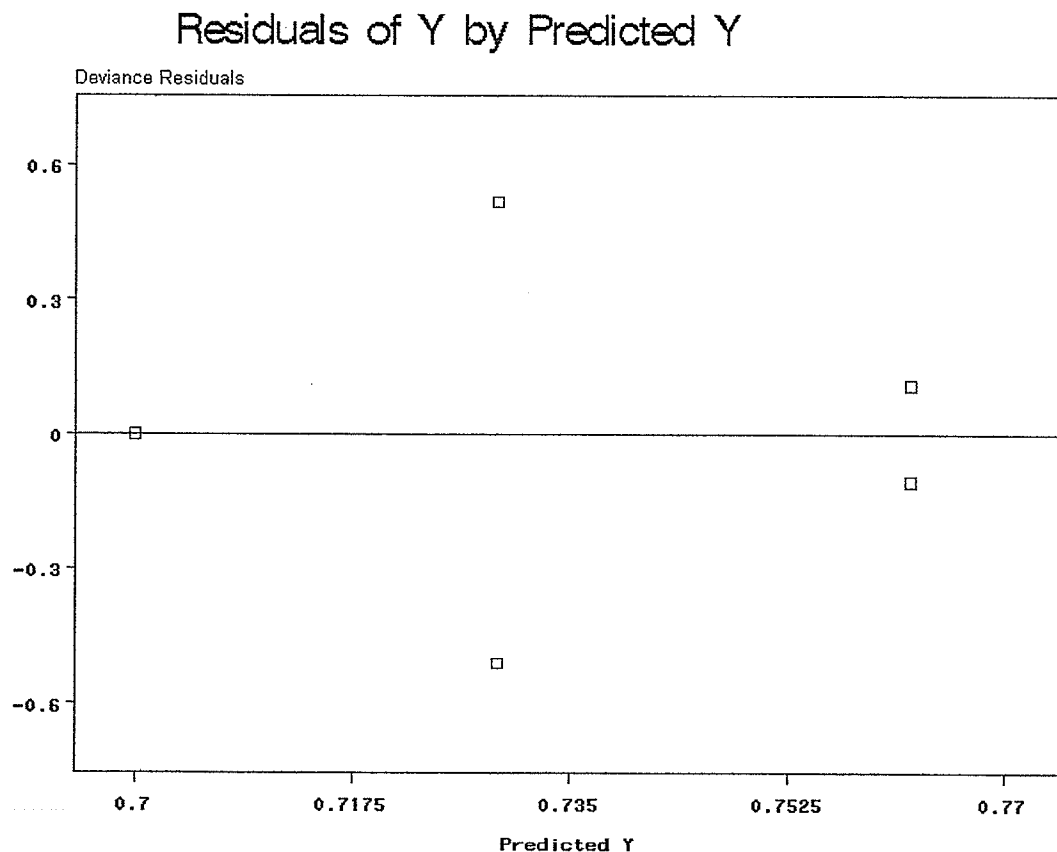
### LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
temp	2	1.72	0.4227

The test statistic is  $\chi^2 = 1.72$  and the corresponding p-value is 0.4772, we again see that the temperature effect is not significant.

Lastly, we will look at a residual plot:

Figure 2.2:



As we can see the residual plot at this stage only contains five data points. Unfortunately there is not enough information to make meaningful judgements.

## ***Chapter 3: Analysis of the Moistures***

### **3.1 INTRODUCTION**

In the previous chapter, we looked at temperature and its impact on the survival of a turtle.

The terms involved in the whole plot model, aside from the error, were fixed effects.

At the split plot level, we introduce a new factor: moisture. The factor moisture has two levels: wet and dry. There are three trays in each incubator. One of the two moisture levels was randomly assigned to each tray such that there were two wet and one dry /or two dry and one wet tray per incubator.

At this time, we want to see how the factor moisture and the interaction of moisture with temperature influence the survival of a turtle.

With the introduction of moisture it is no longer the case that there is one random term i.e. error. Since the trays and incubators are considered random we need to reflect this in the model. An extension of the generalized linear model that takes into consideration random effects is a generalized linear mixed effects model or generalized linear mixed model.

The generalized linear mixed model contains fixed effects as well as at least two random terms, one of which accounts for the model error. The mixed model can be written in the following matrix form:

$$\underline{\eta} = X \underline{\beta} + Z \underline{u}$$

where:

$\underline{\eta}$  is the link function, here a vector of logit means.

X is the design matrix associated with the fixed effects

$\underline{\beta}$  are fixed effect parameters

Z is the design matrix associated with random effects

$\underline{u}$  are random effects  $u \sim (0, G)$ .

Let  $\underline{\varepsilon}$  be an unknown random error vector  $\varepsilon \sim (0, R)$

When averaged over  $\underline{u}$ , the expectation is:  $\underline{\eta} = X \underline{\beta}$ .

The variance –covariance structure is typically represented as  $Var \begin{bmatrix} \underline{u} \\ \underline{\varepsilon} \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$ .

Estimates for G and R can be found using SAS. To analyze a generalized linear mixed model we will use extensions and analogues to the techniques used in the generalized linear model.

We recall that a generalized linear model has observations  $y_1, \dots, y_n$  with means  $\mu_1, \dots, \mu_n$ . The response variable has a distribution which is a member of the exponential family. It has a set of parameters  $\underline{\beta}$  and explanatory variables  $x' = [x_1, \dots, x_p]$  associated with it. There is a link function around which the model is built. The link function connects the mean of the raw data to the linear predictor ( $\eta$ ).

In a generalized linear mixed model the parameters of the model  $\underline{\beta}$ , G and R must be estimated. To estimate the unknown parameters we will extend

the method of maximum likelihood and Fisher scoring as in generalized linear models.

The non – error random effects in the mixed model are seen in the  $Zu$  portion of the model. The effects in  $u$  are not parameters, but random variables. In the theory of mixed models, inference is/can be done on the non – error random effects  $u$ . The type of inference that is done is prediction. The predictor of  $u$  is known as the Best Linear Unbiased Predictor (BLUP). It is found by methods analogous to those used to estimate  $\beta$ .

The next chapter will look in detail at BLUP inference.

For parameter estimation in mixed models a procedure that looks at both fixed and random effects at the same time is needed. PROC MIXED in SAS does that. The MIXED procedure, used when dealing with mixed models, is for normally distributed data. As we can see with the turtle data, the response variable is not normal. This is the reason for using generalized linear model procedures in the previous chapter. To cater to these conditions and the mixed model we use the GLIMMIX macro, a SAS macro written for fitting generalized linear mixed models using PROC MIXED estimation methods.

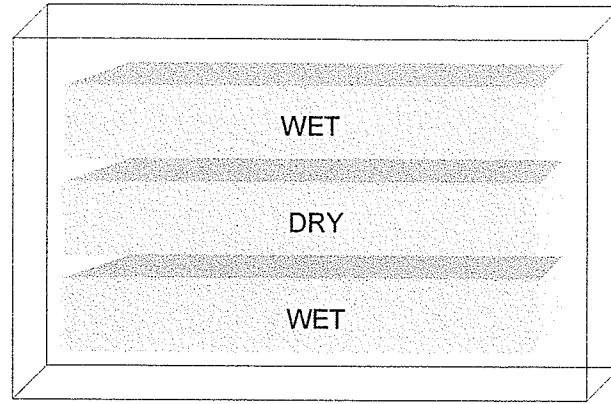
## 3.2 MOISTURE ANALYSIS

At the whole plot level we found that temperature had no effect on the survival of a turtle.

At the next level, the subplot level, we are going to consider the added effect of moisture and interaction of moisture by temperature that might exist.

The following front view of an incubator portrays the subplot level.

Figure 3.1:



We have an incubator at one of three temperatures. In each incubator there are three trays; each tray is randomly assigned to one of two moisture levels, there are one wet and two dry trays or two wet and one dry tray.

In the last chapter we used the logit link to fit the data. Continuing with the same link function at the subplot level we define the generalized linear mixed model for the turtle data as follows:

$$Y_{ijkl} \sim \text{Binomial}(n_{ijkl}, \pi_{ijkl})$$

Let  $Y_{ijkl}$  equal the number of turtle eggs that survived on tray  $l$  in incubator  $j$  at temperature  $i$  and moisture  $k$ . We define  $\pi_{ijkl}$  as the probability that a turtle egg at the  $i$ th temperature in  $j$ th incubator and  $k$ th moisture in  $l$ th tray successfully hatches.  $n_{ijkl}$  denotes the number of turtle eggs at temperature  $i$  in incubator  $j$  randomly assigned to moisture  $k$  and tray  $l$ .

The generalized linear mixed model can be written as:

$$\log\left(\frac{\pi}{1-\pi}\right) = X\underline{\beta} + Z\underline{u}$$

where  $\log\left(\frac{\pi}{1-\pi}\right)$  represents the vector of mean logits at all temperatures, all moistures in all incubators on all trays. At this level we will have a vector of size (15 x 1).

The  $\underline{\beta}$  vector with dimension (6 x 1) includes the parameters needed at

the subplot level.  $\underline{\beta} = \begin{pmatrix} \mu^{**} \\ \tau_1^* \\ \tau_2^* \\ \gamma_1^* \\ (\tau\gamma)_{11}^* \\ (\tau\gamma)_{21}^* \end{pmatrix}$

Fixed effects will be denoted by Greek letters and random effects by Roman letters.

Continuing with the reference cell formulation we define the parameters as:

$\mu^{**}$  is the mean logit at temperature 29.5<sup>0</sup>C and moisture wet

$\tau_1^*$  and  $\tau_2^*$  are previously defined

$\gamma_1^*$  is the mean logit at moisture dry – mean logit at moisture wet

$(\tau\gamma)_{11}^*$  is the mean logit at temperature 21<sup>0</sup>C and moisture dry – mean logit at temperature 21<sup>0</sup>C and moisture wet – mean logit at temperature 29.5<sup>0</sup>C and moisture dry + mean logit at 29.5<sup>0</sup>C and moisture wet.

$(\tau\gamma)_{21}^*$  is the mean logit at temperature 25°C and moisture dry – mean logit at temperature 25°C and moisture wet – mean logit at temperature 29.5°C and moisture dry + mean logit at 29.5°C and moisture wet.

The X matrix has dimensions (15 x 6) and is as follows:

$$X = \begin{pmatrix} 1 & 1 & . & 1 & 1 & . \\ 1 & 1 & . & 1 & 1 & . \\ 1 & 1 & . & . & . & . \\ 1 & 1 & . & 1 & 1 & . \\ 1 & 1 & . & . & . & . \\ 1 & 1 & . & . & . & . \\ \hline 1 & . & 1 & 1 & . & 1 \\ 1 & . & 1 & 1 & . & 1 \\ 1 & . & 1 & . & . & . \\ 1 & . & 1 & 1 & . & 1 \\ 1 & . & 1 & . & . & . \\ 1 & . & 1 & . & . & . \\ \hline 1 & . & . & 1 & . & . \\ 1 & . & . & . & . & . \\ 1 & . & . & . & . & . \\ \hline \underbrace{1}_{\mu^{**}} & \underbrace{.}_{\tau_i^*} & \underbrace{.}_{\gamma_k^*} & \underbrace{.}_{(\tau\gamma)_{ik}^*} & \underbrace{.}_{(\tau\gamma)_{ik}^*} & \underbrace{.}_{(\tau\gamma)_{ik}^*} \end{pmatrix}$$

The next part of the mixed model is the random effects, which can be expressed in the Z and  $\underline{u}$  matrices.

The vector  $\underline{u}$  consists of the non - error random variables in the experiment at the subplot level. The random components at the subplot level include:  $A_{j(i)}$ ; effect of incubator j within temperature i and  $B_{k(j(i))}$  effect of moisture k within incubator j and temperature i.

The Z and  $\underline{u}$  matrices are defined below:

$$Z = \begin{pmatrix} 1 & . & . & . & . & 1 & . & . & . & . & . & . & . & . & . \\ 1 & . & . & . & . & 1 & . & . & . & . & . & . & . & . & . \\ 1 & . & . & . & . & . & 1 & . & . & . & . & . & . & . & . \\ \hline . & 1 & . & . & . & . & . & 1 & . & . & . & . & . & . & . \\ . & . & 1 & . & . & . & . & . & 1 & . & . & . & . & . & . \\ . & . & . & 1 & . & . & . & . & . & 1 & . & . & . & . & . \\ \hline . & . & . & . & 1 & . & . & . & . & . & 1 & . & . & . & . \\ . & . & . & . & . & 1 & . & . & . & . & . & 1 & . & . & . \\ . & . & . & . & . & . & 1 & . & . & . & . & . & 1 & . & . \\ \hline . & . & . & . & . & . & . & . & 1 & . & . & . & . & 1 & . \\ . & . & . & . & . & . & . & . & . & 1 & . & . & . & . & 1 \\ . & . & . & . & . & . & . & . & . & . & 1 & . & . & . & . \\ \hline . & . & . & . & . & . & . & . & . & . & . & . & . & . & 1 \\ \hline \underbrace{\quad \quad \quad}_{A_{j(i)}} & \underbrace{\quad \quad \quad}_{B_{k(j(i))}} \end{pmatrix}$$

$$\underline{u} = \begin{bmatrix} A_{1(1)} \\ A_{2(1)} \\ A_{1(2)} \\ A_{2(2)} \\ A_{1(3)} \\ B_{1(1(1))} \\ B_{2(1(1))} \\ B_{1(2(1))} \\ B_{2(2(1))} \\ B_{1(1(2))} \\ B_{2(1(2))} \\ B_{1(2(2))} \\ B_{2(2(2))} \\ B_{1(1(3))} \\ B_{2(1(3))} \end{bmatrix}$$

The dots represent zeros.

The G – matrix is the variance - covariance matrix for the random components in the mixed model. The matrix is a 15x15 diagonal matrix where the first 5x5 submatrix along the diagonal has the variance associated with incubator within temperature and the remaining 10x10 submatrix along the diagonal has the variance associated with moisture within incubator and temperature.

### 3.2.1 Penalized Quasi - likelihood

The version of the likelihood that SAS uses in generalized linear mixed models is penalized quasi - likelihood. Penalized quasi - likelihood as used in generalized linear mixed models is the analog to quasi - likelihood used in generalized linear models. In this situation (i.e. a generalized linear mixed model) the quasi - likelihood has been augmented by a term involving  $\underline{u}$  (i.e.  $\exp(\underline{u}'G^{-1}\underline{u})$ ). The mixed model relates  $\underline{\beta}$  to  $\underline{\eta}$  by using the expected value of  $\underline{\eta}$  given  $\underline{u}$ . The conditional mean of  $\underline{\eta}$  given  $\underline{u}$  is related to  $\underline{\beta}$  via the equation  $\underline{\eta} = X\underline{\beta} + Z\underline{u}$ . The covariance matrix of  $\underline{u}$ , denoted by  $G$ , depends on the vector of parameters  $\underline{\theta}$ . The variance contains terms involving the variance of  $Z\underline{u}$  and the variance of  $\underline{\varepsilon}$  (i.e.  $Z'GZ + R$ ).

Similarly to section 2.4, we wish to estimate  $\underline{\beta}$  and  $\underline{\theta}$  using maximum likelihood or an approximation to it. We recall in section 2.4 there was difficulty in implementing likelihood methods directly. To estimate  $\underline{\beta}$  and  $\underline{\theta}$  when using the penalized quasi - likelihood, a method of integral approximation is used.

### 3.2.2 Fitting the model

To fit the generalized linear mixed model we use the GLIMMIX macro.

The following selected SAS output was obtained.

First we look at how well the model fits the data:

### GLIMMIX Model Statistics

Description	Value
Deviance	11.5797
Scaled Deviance	10.9453
Pearson Chi-Square	11.6598
Scaled Pearson Chi-Square	11.0211
Extra-Dispersion Scale	1.0580

Similarly, to the previous chapter we will compare the deviance to a  $\chi^2_9$ .

We find the p-value=0.2381 which suggests the model is a good fit.

Solution for Fixed Effects							
Effect	Moist	Temp	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept			0.8473	0.3239	2	2.62	0.1203
temp		21	0.1908	0.3741	2	0.51	0.6608
temp		25	0.4363	0.3821	2	1.14	0.3718
moist	D		0.04256	0.4691	2	0.09	0.9360
moist*temp	D	21	-0.09333	0.5588	2	-0.17	0.8827
moist*temp	D	25	-0.2148	0.5689	2	-0.38	0.7420

The output suggests that with the addition of the moisture effect we still have no significant factors i.e. the above factors have no effect on the survival of a turtle egg.

Estimates for the random effects provide us with variance estimates for the 3 random components in our model.

Covariance Parameter Estimates	
Cov Parm	Estimate
inc(temp)	-0.03564
moist*inc(temp)	0.01459
Residual	1.0580

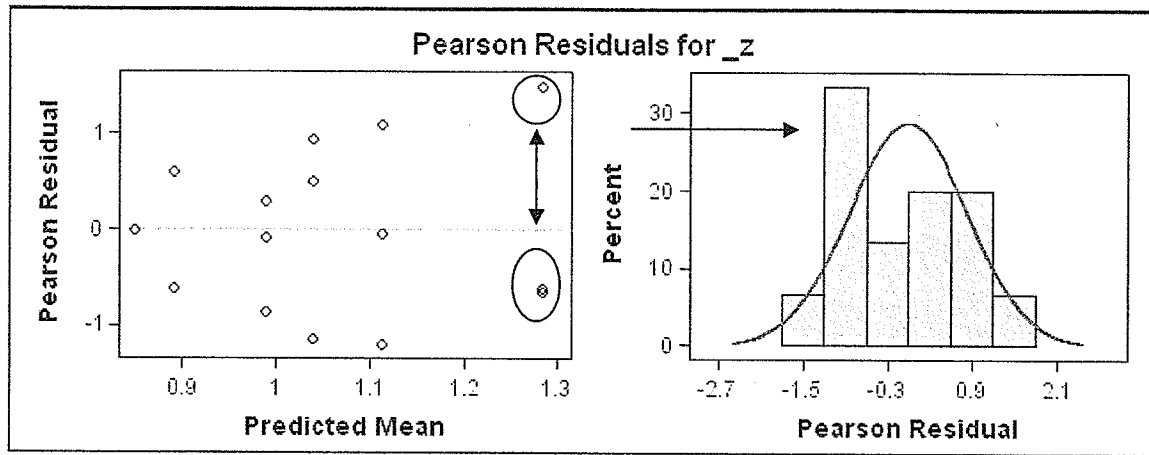
i.e.  $\hat{\sigma}_{inc(temp)}^2 = -0.03564$ ,  $\hat{\sigma}_{moist*inc(temp)}^2 = 0.01459$  and the subplot model error  $\hat{\sigma}^2 = 1.0580$ . The estimate  $\hat{\sigma}_{inc(temp)}^2 = -0.03564$  is negative. We will replace the value by zero i.e.  $\hat{\sigma}_{inc(temp)}^2 = 0$ . This suggests that the variance of the data is less than that predicted by the binomial model.

Next we present the solutions for the random effects.

Solution for Random Effects								
Effect	moist	inc	temp	Estimate	Std Err Pred	DF	t Value	Pr >  t
inc(temp)		1	21	-0.2090	0	5	-Infy	<.0001
inc(temp)		2	21	0.2090	0	5	Infy	<.0001
inc(temp)		1	25	-0.08359	0	5	-Infy	<.0001
inc(temp)		2	25	0.08359	0	5	Infy	<.0001
inc(temp)		1	29	7.93E-16	.	.	.	.
moist*inc(temp)	D	1	21	0.04660	0.1140	5	0.41	0.6997
moist*inc(temp)	D	2	21	-0.04660	0.1140	5	-0.41	0.6997
moist*inc(temp)	W	1	21	0.03899	0.1138	5	0.34	0.7458
moist*inc(temp)	W	2	21	-0.03899	0.1138	5	-0.34	0.7458
moist*inc(temp)	D	1	25	0.04647	0.1146	5	0.41	0.7019
moist*inc(temp)	D	2	25	-0.04647	0.1146	5	-0.41	0.7019
moist*inc(temp)	W	1	25	-0.01224	0.1153	5	-0.11	0.9196
moist*inc(temp)	W	2	25	0.01224	0.1153	5	0.11	0.9196
moist*inc(temp)	D	1	29	-164E-18	0.1208	5	-0.00	1.0000
moist*inc(temp)	W	1	29	-144E-18	0.1208	5	-0.00	1.0000

Lastly we look at a residual plot to see if there are any patterns.

Figure 3.2:



As we can see there are no patterns in the data set. There are a few points in the residual plot that could be considered influential observations.

## ***Chapter 4: Analysis of the Sites***

### **4.1 INTRODUCTION**

In chapter two a detailed look at estimating a fixed effect (i.e. temperature, here) was completed. For this fixed effect analysis of the turtle data, a generalized linear model was used. For the generalized linear model the distribution form of the response variable is known. In the case of the turtle data the response variable followed a binomial distribution. With the distribution known, the likelihood function can be found. Once the likelihood function is found it is then maximized to find the estimates.

In the case of the turtle data it was found that the likelihood function became too complex; as a result Quasi – likelihood was used. Similarly to the maximum likelihood, the quasi-likelihood function is maximized to find the estimates of the parameters.

In the last chapter we introduced random effects terms into the model, denoted by  $\underline{u}$ . By including random effects the model became a generalized linear mixed model. To obtain estimates for the mixed model a penalized quasi – likelihood function was maximized.

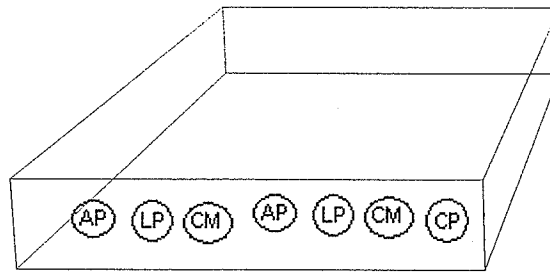
The next stage of the analysis is to look at site effects. We will also predict the survival of a turtle egg in various situations.

## 4.2 SITES ANALYSIS

At the sub-sub-plot level we consider the factor site. The experiment was conducted at four sites in Ontario. With the help of our initial plot in figure 1.4, it can be seen that site had an effect on the survival of a turtle egg.

Figure 4.1 represents a top view of a tray in an incubator.

Figure 4.1:



Considering the factor site the following can be defined:

$$Y_{ijklm} \sim B(n_{ijklm}, \pi_{ijklm})$$

where:  $Y_{ijklm}$  is the number of turtle eggs that survived at temperature  $i$ , in incubator  $j$  at moisture  $k$  on tray  $l$  from site  $m$ .  $n_{ijklm}$  is the total number of turtle eggs at temperature  $i$ , in incubator  $j$  at moisture  $k$  on tray  $l$  from site  $m$ . Finally,  $\pi_{ijklm}$  is the probability that a turtle egg survived at temperature  $i$ , in incubator  $j$  at moisture  $k$  on tray  $l$  from site  $m$ .

Continuing with the logit link, the generalized linear mixed model can be written as:

$$\log\left(\frac{\pi}{1-\pi}\right) = X\underline{\beta} + Z\underline{u}$$

At this time the dimension and terms involved in the mixed model will be considered. The  $\log\left(\frac{\pi}{1-\pi}\right)$  vector is of size (60x1). This takes into account the four sites, three trays per incubator and five incubators.  $\log\left(\frac{\pi}{1-\pi}\right)$  represents

the vector of 60 mean logits. where  $\underline{\pi} = \begin{pmatrix} \pi_{1111} \\ \vdots \\ \pi_{31124} \end{pmatrix}$  ie:  $\pi_{31124}$  represents the mean

logit at temperature 29.5°C, in incubator 1 at moisture dry on tray 2 and from site LP.

There are two ways of considering the factor site; both possibilities will be modeled. In the first way if the locations were predetermined then sites would be considered fixed. On the other hand, if the sites were randomly chosen from a variety of possible sites then sites would be considered random.

First site is considered a fixed factor.

#### 4.2.1 FIXED SITES

Let us analyze the data assuming sites are fixed. The fixed effects are represented by the X matrix and the  $\underline{\beta}$  vector.  $\underline{\beta}$  has dimension (18x1) and includes the parameters from the sub – plot level  $\tau_1^*, \tau_2^*, \gamma_1^*, (\tau\gamma)_{11}^*, (\tau\gamma)_{21}^*$  which were previously defined. In addition the following new parameters are added.

$$\left( \mu^{\dots} \quad \delta_1^* \quad \delta_2^* \quad \delta_3^* \quad (\tau\delta)_{11}^* \quad (\tau\delta)_{12}^* \quad (\tau\delta)_{13}^* \quad (\tau\delta)_{21}^* \quad (\tau\delta)_{22}^* \quad (\tau\delta)_{23}^* \quad (\gamma\delta)_{11}^* \quad (\gamma\delta)_{12}^* \quad (\gamma\delta)_{13}^* \right)$$

Continuing with the reference cell formulation the additional parameters can be defined as follows:

$\mu^{***}$  is the mean logit at temperature 29.5°C, moisture wet and site LP.

$\delta_1^*$  is the mean logit at site AP – mean logit at site LP

$\delta_2^*$  is the mean logit at site CM – mean logit at site LP

$\delta_3^*$  is the mean logit at site CP – mean logit at site LP

$(\tau\delta)_{11}^*$  is the mean logit at temperature 21°C, site AP – mean logit at temperature 21°C, site LP – mean logit at temperature 29.5°C, site AP + mean logit at temperature 29.5°C, site LP.

$(\tau\delta)_{12}^*$  is the mean logit at temperature 21°C, site CM – mean logit at temperature 21°C, site LP – mean logit at temperature 29.5°C, site CM + mean logit at temperature 29.5°C, site LP.

$(\tau\delta)_{13}^*$  is the mean logit at temperature 21°C, site CP – mean logit at temperature 21°C, site LP – mean logit at temperature 29.5°C, site CP + mean logit at temperature 29.5°C, site LP.

$(\tau\delta)_{21}^*$  is the mean logit at temperature 25°C, site AP – mean logit at temperature 25°C, site LP – mean logit at temperature 29.5°C, site AP + mean logit at temperature 29.5°C, site LP.

$(\tau\delta)_{22}^*$  is the mean logit at temperature 25°C, site CM – mean logit at temperature 25°C, site LP – mean logit at temperature 29.5°C, site CM + mean logit at temperature 29.5°C, site LP.

$(\tau\delta)_{23}^*$  is the mean logit at temperature 25°C, site CP – mean logit at temperature 25°C, site LP – mean logit at temperature 29.5°C, site CP + mean logit at temperature 29.5°C, site LP.

$(\gamma\delta)_{11}^*$  is the mean logit at moisture dry, site AP – mean logit at moisture dry, site LP – mean logit at moisture wet, site AP + mean logit at moisture wet, site LP.

$(\gamma\delta)_{12}^*$  is the mean logit at moisture dry, site CM – mean logit at moisture dry, site LP – mean logit at moisture wet, site CM + mean logit at moisture wet, site LP.

$(\gamma\delta)_{13}^*$  is the mean logit at moisture dry, site CP – mean logit at moisture dry, site LP – mean logit at moisture wet, site CP + mean logit at moisture wet, site LP.

The design matrix, X is of size (60x18). We present the part for the first temperature.

1	1	.	1	1	.	1	.	.	1	.	.	.	.	1	.	.
1	1	.	1	1	.	1	.	.	1	.	.	.	.	1	.	.
1	1	.	1	1	.	.	1	.	.	1	.	.	.	.	1	.
1	1	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.
1	1	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.
1	1	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.
1	1	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.
1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	1	.	1	1	.	1	.	.	1	.	.	.	.	1	.	.
1	1	.	1	1	.	.	1	.	.	1	.	.	.	.	1	.
1	1	.	1	1	.	.	.	1	.	.	1	.	.	.	.	1
1	1	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.
1	1	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.
1	1	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.
1	1	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.
1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	1	.	1	1	.	1	.	.	1	.	.	.	.	1	.	.
1	1	.	1	1	.	.	1	.	.	1	.	.	.	.	1	.
1	1	.	1	1	.	.	.	1	.	.	1	.	.	.	.	1
1	1	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.
1	1	.	.	.	.	1	.	.	1	.	.	.	.	.	.	.
1	1	.	.	.	.	.	1	.	.	1	.	.	.	.	.	.
1	1	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.
1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

As can be seen a pattern of sub-matrices arise. The above matrix can be written as:

$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$
$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$0_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$
$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$
$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$0_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$
$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$
$J_{(4 \times 2)}$	$0_{(4 \times 1)}$	$0_{(4 \times 2)}$	$0_{(4 \times 1)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$I_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$	$0_{(3 \times 3)}$ $0_{(1 \times 3)}$

The identity matrix (denoted by I) is repeated every three lines to show the different sites followed by a line of zeros (represented by 0) to denote the reference site. The J sub-matrix represents a matrix of 1's. The matrix size is written in brackets.

The random components of the mixed model are in Z and  $\underline{u}$ . The  $\underline{u}$  vector has dimension (20x1) and includes the following random effects:

$A_{j(i)}$  incubator within temperature. (There are two incubators at each of three temperature levels minus the one incubator that malfunctioned at temperature 29.5°C.)

$B_{I(j(i) \times k)}$  tray nested within incubator within temperature by moisture. For some reason SAS would not accept this formulation. The random effect was rewritten as  $B_{I(i \times j \times k)}$  i.e. tray nested within incubator, temperature and moisture. A portion of the vector can be written as follows:

$$\left( A_{1(1)} \quad A_{2(1)} \quad A_{1(2)} \quad A_{2(2)} \quad A_{1(3)} \quad B_{1(111)} \quad B_{1(112)} \quad B_{2(111)} \quad B_{1(122)} \quad \dots B_{2(311)} \right)$$

The design matrix  $Z$  of the random effects is of size (60x20). The portion of  $Z$  for temperature 21<sup>0</sup>C can be written as:

The figure shows a 16x16 grid of points. A 4x4 subgrid of points is marked with '1'. The subgrid is labeled  $A_j(l)$  and  $B_l(i \times j \times k)$ .

The portion of Z for temperature 21°C can be rewritten as a set of sub matrices.

$J_{(12 \times 1)} \quad O_{(12 \times 1)}$	$O_{(24 \times 3)}$	$J_{(4 \times 1)} \quad O_{(4 \times 5)}$	$O_{(4 \times 9)}$
		$O_{(4 \times 1)} \quad J_{(4 \times 1)} \quad O_{(4 \times 4)}$	$O_{(4 \times 9)}$
		$O_{(4 \times 2)} \quad J_{(4 \times 1)} \quad O_{(4 \times 3)}$	$O_{(4 \times 9)}$
$O_{(4 \times 3)} \quad J_{(4 \times 1)} \quad O_{(4 \times 2)}$		$O_{(4 \times 9)}$	
$O_{(4 \times 4)} \quad J_{(4 \times 1)} \quad O_{(4 \times 1)}$		$O_{(4 \times 9)}$	
$O_{(4 \times 5)} \quad J_{(4 \times 1)}$		$O_{(4 \times 9)}$	
$O_{(12 \times 1)} \quad J_{(12 \times 1)}$			

As we can see there is a pattern developing for temperature 21°C. A similar pattern can be seen for temperature 25°C and 29.5°C.

At this time the generalized linear mixed model for the turtle data can be stated as:

$$E \left[ \log \frac{Y_{ijklm}}{1 - Y_{ijklm}} \middle| A_{j(i)}, B_{l(i \times j \times k)} \right] = \mu^{***} + \tau_i^* + A_{j(i)} + \gamma_k^* + (\tau\gamma)_{ik}^* + B_{l(i \times j \times k)} + \delta_m^* + (\tau\delta)_{im}^* + (\gamma\delta)_{km}^*$$

Where:

$k = 1, 2$

$A_{j(i)} \sim (0, \sigma_a^2)$

$i = 1, 2, 3$

$l = 1, 2, 3, \dots n_{ijk}$

$B_{l(i \times j \times k)} \sim (0, \sigma_b^2)$

$j = 1, 2, \dots n_i$

$m = 1, 2, 3, 4$

Using the SAS GLIMMIX macro to fit the model the following selected SAS output was obtained to analyze the generalized linear mixed model.

First, we determine how well the model fits the data. To do so the deviance value given in the output is used:

### ***GLIMMIX Model Statistics***

<b>Description</b>	<b>Value</b>
Deviance	22.4526
Scaled Deviance	43.1196
Pearson Chi-Square	19.3548
Scaled Pearson Chi-Square	37.1703
Extra-Dispersion Scale	0.5207

The deviance value is compared to a  $\chi^2$  with n-p degrees of freedom. We find

$\Pr(\chi^2_{42} > 22.4536) = 0.9942$  suggesting the model does fit the data.

Next, we estimate the three variance components which are called covariance parameter estimates by SAS. These estimates correspond to the random sources of variation. The output is:

<b>Covariance Parameter Estimates</b>	
<b>Cov Parm</b>	<b>Estimate</b>
<b>inc(temp)</b>	-0.06109
<b>tray(moist*inc*temp)</b>	0.1751
<b>Residual</b>	0.5207

The covariance parameter inc(temp) is the estimate for  $\sigma_a^2$ . It represents the variation (or error) associated with the different incubators used for the turtle experiment. The estimate was found to be -0.06109. The estimates are in terms of the logit scale.

Estimates of the fixed effect parameters in the model are as follows: The estimates are given in terms of the logit link function. We recall that SAS uses reference cell models.

Solution for Fixed Effects								
Effect	site	moist	temp	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept				2.7300	0.7909	2	3.45	0.0746
temp			21	-0.2736	0.7908	2	-0.35	0.7623
temp			25	1.5245	1.0191	2	1.50	0.2733
moist		D		-0.2139	0.8187	7	-0.26	0.8014
moist*temp		D	21	0.09979	0.7197	7	0.14	0.8936
moist*temp		D	25	-0.1508	0.7279	7	-0.21	0.8418
site	AP			-0.9880	0.7981	33	-1.24	0.2245
site	CM			-2.7860	0.7555	33	-3.69	0.0008
site	CP			-2.4485	0.7454	33	-3.28	0.0024
site*temp	AP		21	1.0155	0.7823	33	1.30	0.2033
site*temp	AP		25	-0.8142	1.0045	33	-0.81	0.4234
site*temp	CM		21	1.0510	0.7307	33	1.44	0.1598
site*temp	CM		25	-0.3191	0.9768	33	-0.33	0.7460
site*temp	CP		21	-0.5383	0.7192	33	-0.75	0.4595

Solution for Fixed Effects								
Effect	site	moist	temp	Estimate	Standard Error	DF	t Value	Pr >  t
site*temp	CP		25	-1.8878	0.9616	33	-1.96	0.0581
site*moist	AP	D		-0.07428	0.6740	33	-0.11	0.9129
site*moist	CM	D		0.05259	0.6250	33	0.08	0.9334
site*moist	CP	D		0.2843	0.6147	33	0.46	0.6468

The estimates of the temperature and moisture main effects and temperature by moisture interaction are consistent with the last section; therefore they are not discussed here.

For the site effect, the SAS output indicates differences, similarly to the results found in figure 1.4. Specifically, we find that site CM (Cranberry Marsh) (p-value = 0.0008) and CP (Cootes' Paradise) (p-value = 0.0024) are significantly lower than LP (Big Creek Marsh) while AP (Algonquin Park) is not significantly different than LP.

Next, we will look at the interactions involving site.

For the site by moisture interaction: In figure 1.6 the interaction plots showed no interaction between these two factors. The output above agrees with this.

For the site by temperature interaction: In figure 1.7 the interaction plots showed some differences. The SAS output shows the only interaction that approaches significance is site CP at temperature 25°C  $(\tau\delta)_{23}^*$  it has p-value of 0.0581 which is significant at 10% but not at 5%. Recall, that the site by

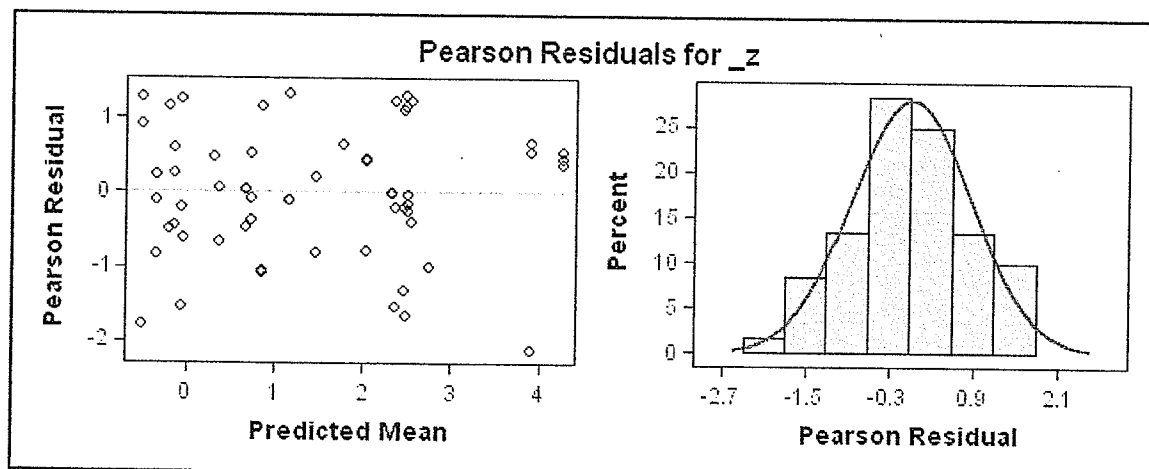
temperature interaction is a contrast of four logits,  $(\tau\delta)_{23}^*$  is the mean logit at temperature 25°C, site CP – mean logit at temperature 25°C, site LP – mean logit at temperature 29.5°C, site CP + mean logit at temperature 29.5°C, site LP.

Using our model we can find estimated probabilities that a turtle egg will survive given various conditions. For example the following SAS output gives the estimate probability of a turtle egg surviving at temperature 21, moisture dry and from site AP as 0.9085. Similarly we find the estimated probability that a turtle egg will survive temperature 21, moisture dry and from site CM as 0.6592.

Estimates	
Label	$\hat{\pi}_{1 \cdot 1 \cdot 1}$
21/D/AP/	0.9085
21/D/CM/	0.6592
25/D/AP/	0.8822

Lastly we will look at the residual plot:

Figure 4.2:



As we can see there are no patterns in the residual plot.

#### 4.2.2 RANDOM SITES

Alternatively to the previous section we could consider sites as being random.

First, we consider the fixed effects represented in the X matrix and  $\underline{\beta}$  vector. The  $\underline{\beta}$  vector has dimension (6x1) and includes the parameters

$\left( \mu^{**}, \tau_1^*, \tau_2^*, \gamma_1^*, (\tau\gamma)_{11}^*, (\tau\gamma)_{21}^* \right)'$  from the sub – plot level which were previously defined.

The design matrix, X is of size (60x6). The following portion represents the fixed effects at temperature 21°C.

1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	1	1	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.
1	1	.	.	.	.

$\mu^{**}$

$\tau_j^*$

$\gamma_R^*$

$(\tau\gamma)_{Rk}^*$

The X matrix can be rewritten as a set of submatrices:

$J_{4 \times 2}$	$O_{4 \times 1}$	$J_{4 \times 2}$	$O_{4 \times 1}$
$J_{4 \times 2}$	$O_{4 \times 1}$	$O_{4 \times 2}$	$O_{4 \times 1}$
$J_{4 \times 2}$	$O_{4 \times 1}$	$J_{4 \times 2}$	$O_{4 \times 1}$
$J_{4 \times 2}$	$O_{4 \times 1}$	$O_{4 \times 2}$	$O_{4 \times 1}$
$J_{4 \times 2}$	$O_{4 \times 1}$	$J_{4 \times 2}$	$O_{4 \times 1}$
$J_{4 \times 2}$	$O_{4 \times 1}$	$O_{4 \times 2}$	$O_{4 \times 1}$

The random portion seen in the  $\underline{u}$  matrix has dimension (44x1) and includes the following terms:

$A_{f(l)}$  incubator within temperature

$B_{l(i \times j \times k)}$  tray nested within incubator, temperature and moisture.

$S_m$  main effect of site.

$(\tau S)_{im}$  temperature by site interaction.

$(\gamma S)_{km}$  moisture by site interaction.

The random design matrix Z is of size (60x44). Because the matrix is so big it is efficient to write it as a pattern of submatrices. The portion for temperature 21°C can be written as follows:

$$\left\{ \begin{array}{cccccc} J_{4 \times 1} & O_{4 \times 4} & J_{4 \times 1} & O_{4 \times 14} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} \\ J_{4 \times 1} & O_{4 \times 4} & O_{4 \times 1} & J_{4 \times 1} & O_{4 \times 13} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} \\ J_{4 \times 1} & O_{4 \times 4} & O_{4 \times 2} & J_{4 \times 1} & O_{4 \times 12} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} \\ O_{4 \times 1} & J_{4 \times 1} & O_{4 \times 3} & O_{4 \times 3} & J_{4 \times 1} & O_{4 \times 11} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} \\ O_{4 \times 1} & J_{4 \times 1} & O_{4 \times 3} & O_{4 \times 4} & J_{4 \times 1} & O_{4 \times 10} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} \\ O_{4 \times 1} & J_{4 \times 1} & O_{4 \times 3} & O_{4 \times 5} & J_{4 \times 1} & O_{4 \times 9} & I_{4 \times 4} & I_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & O_{4 \times 4} & I_{4 \times 4} \\ & A_{j(i)} & B_{l(i \times j \times k)} & S_m & (\tau S)_{im} & (\gamma S)_{km} \end{array} \right\}$$

The generalized linear mixed model for the turtle data can be stated as:

$$E \left[ \log \frac{Y_{ijklm}}{1 - Y_{ijklm}} \middle| A_{j(i)}, B_{l(i \times j \times k)}, S_m, (\tau S)_{im}, (\gamma S)_{km} \right] = \mu^{**} + \tau_i^* + A_{j(i)} + \gamma_k^* + (\tau \gamma)_{ik}^* + B_{l(i \times j \times k)} + S_m + (\tau S)_{im} + (\gamma S)_{km}$$

Where:

$$\begin{array}{ll} i = 1, 2, 3 & A_{j(i)} \sim (0, \sigma_a^2) \\ j = 1, 2, \dots, n_i & B_{l(i \times j \times k)} \sim (0, \sigma_b^2) \\ k = 1, 2 & S_m \sim (0, \sigma_s^2) \\ l = 1, 2, 3, \dots, n_{ijk} & (\tau S)_{im} \sim (0, \sigma_T^2) \\ m = 1, 2, 3, 4 & (\gamma S)_{km} \sim (0, \sigma_U^2) \end{array}$$

Using the SAS GLIMMIX macro to fit the model the following selected output was obtained.

First, determine how well the model fits the data.

### ***GLIMMIX Model Statistics***

<b>Description</b>	<b>Value</b>
Deviance	24.6602
Scaled Deviance	52.6238
Pearson Chi-Square	20.2377
Scaled Pearson Chi-Square	43.1864
Extra-Dispersion Scale	0.4686

Using the deviance value given by SAS and comparing it to a chi-square distribution  $\Pr(\chi^2_{54} > 24.6602) = 0.9998$  we find the model does fit the data.

Next, the six variance or covariance parameters are estimated. The output is:

<b>Covariance Parameter Estimates</b>	
<b>Cov Parm</b>	<b>Estimate</b>
site	1.7708
site*temp	0.2134
site*moist	-0.01450
inc(temp)	-0.05733
tray(moist*inc*temp)	0.1725
Residual	0.4686

In terms of the turtle data these covariance parameters estimate the variation associated with each of the random components seen at the sub-sub plot level.

The SAS output for the fixed effects was found to be consistent with the previous fixed effects output; therefore it is not displayed here.

The following SAS output was found for the random effects in the mixed model.

Solution for Random Effects										
Effect	site	moist	tray	inc	temp	Estimate	Std Err Pred	DF	t Value	Pr >  t
site	AP					0.6827	0.7134	33	0.96	0.3455
site	CM					-0.7742	0.71	33	-1.09	0.2834
site	CP					-1.3635	0.7089	33	-1.92	0.0631
site	LP					1.4551	0.7252	33	2.01	0.0531
site*temp	AP				21	0.3285	0.3649	33	0.90	0.3745
site*temp	AP				25	-0.08889	0.3664	33	-0.24	0.8098
site*temp	AP				29.5	-0.1574	0.3704	33	-0.42	0.6736
site*temp	CM				21	0.1705	0.3556	33	0.48	0.6347
site*temp	CM				25	0.09084	0.3585	33	0.25	0.8015
site*temp	CM				29.5	-0.3547	0.3638	33	-0.97	0.3367
site*temp	CP				21	-0.2927	0.3535	33	-0.83	0.4137
site*temp	CP				25	-0.3613	0.355	33	-1.02	0.3162
site*temp	CP				29.5	0.4897	0.361	33	1.36	0.1841
site*temp	LP				21	-0.2064	0.3807	33	-0.54	0.5914
site*temp	LP				25	0.3594	0.3945	33	0.91	0.3689

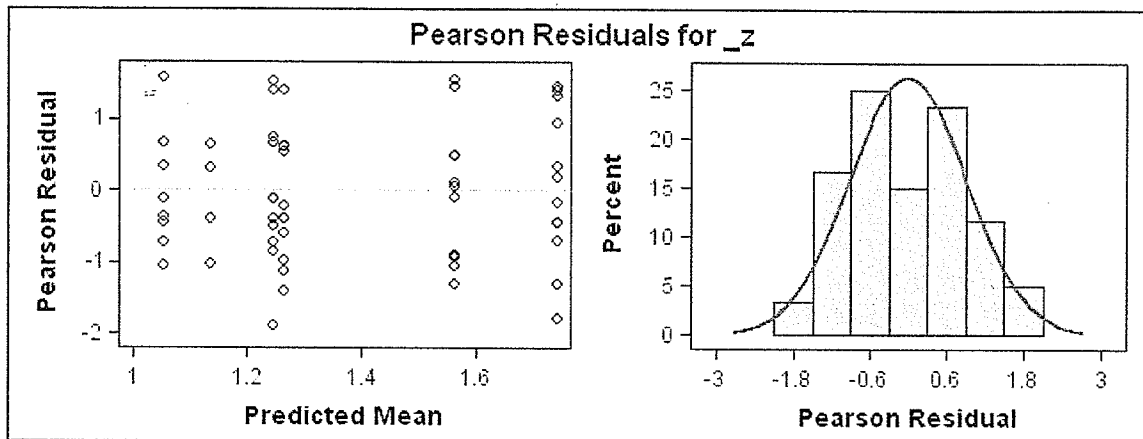
Solution for Random Effects										
Effect	site	moist	tray	inc	temp	Estimate	Std Err Pred	DF	t Value	Pr >  t
site*temp	LP				29.5	0.02232	0.3932	33	0.06	0.9551
site*moist	AP	D				0.03717	0	33	Infty	<.0001
site*moist	AP	W				-0.04276	0	33	Infty	<.0001
site*moist	CM	D				0.0449	0	33	Infty	<.0001
site*moist	CM	W				-0.03856	0	33	Infty	<.0001
site*moist	CP	D				-0.08271	0	33	Infty	<.0001
site*moist	CP	W				0.09388	0	33	Infty	<.0001
site*moist	LP	D				0.000642	0	33	Infty	<.0001
site*moist	LP	W				-0.01256	0	33	Infty	<.0001
inc(temp)				1	21	-0.2599	0	33	Infty	<.0001
inc(temp)				2	21	0.2599	0	33	Infty	<.0001
inc(temp)				1	25	-0.1424	0	33	Infty	<.0001
inc(temp)				2	25	0.1424	0	33	Infty	<.0001
inc(temp)				1	29.5	3.99E-15	.	.	.	.
tray(moist*inc*temp)		D	1	1	21	0.2827	0.2791	33	1.01	0.3184
tray(moist*inc*temp)		D	2	1	21	0.1355	0.2776	33	0.49	0.6286
tray(moist*inc*temp)		D	1	1	25	0.4317	0.2867	33	1.51	0.1417
tray(moist*inc*temp)		D	2	1	25	0.05952	0.2811	33	0.21	0.8336
tray(moist*inc*temp)		D	1	1	29.5	-0.167	0.3332	33	-0.50	0.6196
tray(moist*inc*temp)		D	2	1	29.5	0.167	0.3332	33	0.50	0.6196

Solution for Random Effects										
Effect	site	moist	tray	inc	temp	Estimate	Std Err Pred	DF	t Value	Pr >  t
tray(moist*inc*temp)		D	1	2	21	-0.4182	0.1736	33	-2.41	0.0217
tray(moist*inc*temp)		D	1	2	25	-0.4912	0.186	33	-2.64	0.0125
tray(moist*inc*temp)		W	1	1	21	0.3638	0.181	33	2.01	0.0527
tray(moist*inc*temp)		W	1	1	25	-0.06275	0.1948	33	-0.32	0.7494
tray(moist*inc*temp)		W	1	1	29.5	-1.04E-14	0.4153	33	0.00	1.0000
tray(moist*inc*temp)		W	1	2	21	-0.5468	0.2782	33	-1.97	0.0578
tray(moist*inc*temp)		W	2	2	21	0.183	0.2835	33	0.65	0.5230
tray(moist*inc*temp)		W	1	2	25	-0.2738	0.2831	33	-0.97	0.3406
tray(moist*inc*temp)		W	2	2	25	0.3365	0.2936	33	1.15	0.2600

From the above SAS output there are a few points to consider. First there are a few combinations of interaction effects that have an estimated variance component that is less than zero and as a result the terms involved in these interactions have infinite t-values. The zero standard error is a consequence of a negative variance estimate. This happens to the incubator within temperature and the site by moisture variances. In view of that, the site by moisture interactions and incubator within temperature effects have no meaningful influence on the survival of a turtle egg.

Next we will look at the residual plot.

Figure 4.3:



As we can see the residual plot indicates a good fit.

Generally speaking, in the case of the turtle data when site is considered either random or fixed it is known to affect the survival of a turtle egg. This can be seen in figure 1.4. Next we will attempt to predict the survival of a turtle egg using the generalized linear mixed model.

### 4.3 Best Linear Unbiased Predictors

In the case of this generalized linear mixed model and according to the book Contemporary Statistical Models for the Plant and Soil Sciences [10]; there are two main reasons for incorporating random factors into models:

- 1) 'They accommodate correlated data' [10]. For example the turtle eggs were collected from four Ontario sites. You would expect that each turtle egg would have been laid and collected in similar conditions and around the same time suggesting correlation amongst the eggs from a site.

- 2) 'Allows for broader inference' [10] – this is done by making inferences to a larger population. For example by allowing sites to be random we do not have to draw conclusions specific to the sites chosen for the experiment but our inferences can be made for the overall site effect on the survival of a turtle egg.

To estimate random effects in a mixed model we use **best linear unbiased prediction, or BLUP**.

In statistics, when we use the term predictor (or prediction) we commonly associate it with the outcome of **future** events. In our context, when we refer to the term prediction we will refer to it as the **estimation** of random variables.

Maximizing the penalized quasi-likelihood for  $\underline{\beta}$  and  $\underline{u}$  gives us point predictors for  $\underline{u}$  and point estimates for  $\underline{\beta}$ . Best linear unbiased prediction, or BLUP can be considered in two ways:

- I. Broad Inference or population wide
- II. Narrow Inference or subject-specific

The assumptions needed for BLUP are that  $V = \text{var}(\underline{\varepsilon})$ ,  $\text{cov}(\underline{u}, \underline{\varepsilon})$  and  $\text{var}(\underline{u})$  are known matrices.

Next, we will calculate BLUPs in a few cases using the turtle data.

The factor site will be considered as a random effect in this section.

Similar calculations can be done using site as a fixed effect.

The model is computed on the logit scale. To interpret the value in terms of the raw data the estimates can be transformed by applying the inverse link

function. The link is  $\log\left(\frac{\pi}{1-\pi}\right)$  the predicted value is obtained by using

$$\hat{\pi} = \frac{e^{X\hat{\beta} + Z\hat{u}}}{1 + e^{X\hat{\beta} + Z\hat{u}}}. \text{ The two types of BLUPs can be seen in the following examples.}$$

### 4.3.1 – BROAD INFERENCE

Broad inference or population - wide inference estimates the (fixed) effects of temperature and moisture while averaging over all possible sites, incubators and trays to predict the overall survival of a turtle egg.

Example 1: Considering site as a random factor. Using temperature = 21°C (i=1) and moisture = dry (k=1) we write the model and calculate the predictor as follows:

The expected value is:

$$E\left[\log\frac{Y_{1j1lm}}{1-Y_{1j1lm}} \middle| A_{j(i)}, B_{l(ixjxk)}\right] = \mu^{**} + \tau_1^* + \gamma_1^* + (\tau\gamma)_{11}^* \text{ for any } j, l \text{ and } m$$

Using the estimates found under 'Solutions for fixed effects' from the SAS

output,  $\log\frac{\hat{\pi}_{1j1lm}}{1-\hat{\pi}_{1j1lm}} = 1.26043$  with standard error 0.7358. Recall that the

expected value of the generalized linear model is  $X\hat{\beta}$ . Because this point estimate is given in terms of the link function we transform backwards to get

$$\hat{\pi}_{1j1lm} = \frac{e^{X\hat{\beta}}}{1 + e^{X\hat{\beta}}} \text{ i.e. the predicted probability that a turtle egg will survive from any}$$

site, tray and incubator. The predicted probability is 0.7791. In SAS the predicted standard error is given on a logit scale. To approximate the standard

error in the original scale we can use the delta method. The delta method is used to obtain variances of transformed estimators. In general we have an estimator  $\hat{\theta}$  and a known function  $f(\theta)$ . In our case the known function is the inverse of the logit  $\left( \frac{e^{\theta}}{1+e^{\theta}} \right)$ . The function evaluated at the estimate can be

approximated by a Taylor series  $f(\hat{\theta}) \doteq f(\theta) + (\hat{\theta} - \theta) \frac{df}{d\theta} \Big|_{\theta=\hat{\theta}}$ . In our case we find the variance of  $\hat{\pi}$  to be:  $\text{var}(\hat{\pi}) \doteq \text{var}[\text{logit}(\hat{\pi})] \times \hat{\pi}(1-\hat{\pi})$ . Thus the standard error for the broad inference is approximately 0.3559.

### 4.3.2 NARROW INFERENCE

The second type of BLUP to consider is narrow or subject specific inference. Narrow inference looks at predicting functions that limit inference to the specific random effects observed.

Example 1: Using temperature = 21 ( $i=1$ ) and moisture = dry ( $k=1$ ) the BLUP is taken from SAS. The appropriate quantity to predict is:

$$E \left[ \log \frac{Y_{1j1lm}}{1 - Y_{1j1lm}} \mid A_{j(i)}, B_{l(ixjxk)}, S_m, (\tau S)_{1m}, (\gamma S)_{1m} \right] = \mu^{***} + \tau_1^* + \frac{1}{2} \sum_{j=1}^2 A_{j(i)} + \gamma_1^* + (\tau\gamma)_{11}^* \\ + \frac{1}{3} \sum_{j=1}^2 \sum_{l=1}^{n_{ijk}} B_{l(1xjx1)} + \frac{1}{4} \sum_{m=1}^4 S_m + \frac{1}{4} \sum_{m=1}^4 (\tau S)_{1m} + \frac{1}{4} \sum_{m=1}^4 (\gamma S)_{1m}$$

Similarly to the broad inference, the BLUP is  $\frac{\hat{\pi}_{1j1lm}}{1 - \hat{\pi}_{1j1lm}} = 1.2604$  but now with

standard error 0.1420. Because this BLUP is given in terms of the link function we will apply the inverse link function to find the predicted probability. The

predicted probability that a turtle egg will survive at temperature 21°C and moisture dry from the given trays, incubators and sites used in the experiment is 0.7791. Applying the delta method the corresponding predicted standard error is 0.1563.

As we can see the point estimates for the broad and narrow inference spaces are the same. The difference between broad and narrow inference shows up in the standard error. The transformed standard error for broad inference was calculated as 0.3559 and for narrow inference 0.1563. As we can see, broad inference has a larger standard error than narrow.

The following SAS output was obtained for several BLUPs.

***Sites Random:***

<b>Estimates: Logit Scale</b>					
<b>Label</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>DF</b>	<b>t Value</b>	<b>Pr &gt;  t </b>
<b>21/D/ BLUP BROAD</b>	1.2604	0.7358	7	1.71	0.1305
<b>21/D/ BLUP NARROW</b>	1.2604	0.1420	7	8.88	<.0001
<b>25/D/ BLUP BROAD</b>	1.5593	0.7410	7	2.10	0.0734
<b>25/D/ BLUP NARROW</b>	1.5593	0.1667	7	9.35	<.0001

The label 21/D/BLUP BROAD is interpreted as the broad BLUP for temperature 21°C and moisture dry.

Using the same model we can predict the probability that a turtle egg will survive in each site. The following SAS output shows just that.

Estimates: Logit Scale					
Label	Estimate	Standard Error	DF	t Value	Pr >  t
21/D/AP BLUP NARROW	2.3088	0.2509	7	9.20	<.0001
21/D/CM BLUP NARROW	0.7016	0.1814	7	3.87	0.0061
21/D/CP BLUP NARROW	-0.4785	0.1734	7	-2.76	0.0281
21/D/LP BLUP NARROW	2.5097	0.3005	7	8.35	<.0001

Using our estimates we can apply the inverse link function to find the predicted probabilities. The predicted probability that a turtle egg will survive at temperature 21 ( $i=1$ ), moisture dry ( $k=1$ ) at site AP ( $m=1$ ) is 0.9096. Applying the delta method the corresponding predicted standard error is 0.1436. The following table summarizes the predicted probabilities and standard errors for all sites.

Estimates: Probabilities					
Label	$\hat{\pi}_{1,1,m}$	Predicted Standard Error	DF	t Value	Pr >  t
21/D/AP BLUP NARROW	0.9096	0.1436	7	9.20	<.0001
21/D/CM BLUP NARROW	0.6685	0.2005	7	3.87	0.0061
21/D/CP BLUP NARROW	0.3826	0.2026	7	-2.76	0.0281
21/D/LP BLUP NARROW	0.9248	0.1446	7	8.35	<.0001

The last factor the researcher is interested in is the amount of variation between clutches. We recall that 21 clutches were collected from four sites. Of

those four sites we saw that sites CM and CP were significantly lower than LP. We also saw that the interaction of temperature 25 and site CP was significant at the 10% level.

At this stage we are not able to use our generalized linear mixed model to analyze the clutches. According to Searle and McCullough penalized quasi likelihood 'has not been found to work well in practice, especially for binary data in small clusters' [8]. This can be considered a limitation of generalized linear mixed models.

## ***Chapter 5: Summary***

### **5.1 SUMMARY**

In chapter 1 we made initial main effect and interaction plots showing the proportion of turtle eggs that survived. From these plots we found that the main effect of site and the interaction plot of temperature by site were the only two plots to show some differences in the proportion of turtle eggs that survived.

From the formal analysis, we found the only main effect to be significant was site. We considered site as a fixed effect and as a random effect.

As a fixed effect we found site CM and CP were significantly lower than the reference cell LP while AP showed no difference.

From the interaction plots the only interaction to show some differences was the site by temperature. In fact the SAS output indicated the only interaction term to approach significance was site CP at temperature 25°C.

When sites were considered random we were able to use the estimates from the output to calculate a set of predictors. Using these predictors in the future we can help determine what set of conditions appear to give the highest proportion of turtles surviving.

As a result of our analysis we found our initial plots and formal analysis to agree in their findings.

To analyze the data we used a newly developed technique called Generalized linear models. The dataset used for our analysis was 15 years old. This research shows how to analyze data using this newly developed technique. Research in this area is still ongoing, new developments are continuing.

The results of the experiment show that the factor site cannot be overlooked when examining whether or not a turtle egg survives. A suggested next step for the researcher would be to investigate the environmental differences in these locations ie: were the sites located near cities or were they located in more rural area? Is there a problem with pollution in the sites? etc.

# Appendix A

## CHAPTER 1: Proportion Survived and Other Descriptive Statistics

				Proportion		
Site	Temp	Moist	Alive	Total	Survived	StdError
AP			154	173	0.890	0.024
CM			84	130	0.646	0.042
CP			70	151	0.464	0.041
LP			134	142	0.944	0.019
	21		175	239	0.732	0.029
	25		183	238	0.769	0.027
	29.5		84	119	0.706	0.042
AP	21		64	70	0.914	0.033
CM	21		35	53	0.660	0.065
CP	21		23	58	0.397	0.064
LP	21		53	58	0.914	0.037
AP	25		62	69	0.899	0.036
CM	25		37	51	0.725	0.062
CP	25		29	62	0.468	0.063
LP	25		55	56	0.982	0.018
AP	29.5		28	34	0.824	0.065
CM	29.5		12	26	0.462	0.098
CP	29.5		18	31	0.581	0.089
LP	29.5		26	28	0.929	0.049
		D	232	319	0.727	0.025
		W	210	277	0.758	0.026
AP		D	79	91	0.868	0.035
CM		D	42	69	0.609	0.059
CP		D	37	80	0.463	0.056
LP		D	74	79	0.937	0.027
AP		W	75	82	0.915	0.031
CM		W	42	61	0.689	0.059
CP		W	33	71	0.465	0.059
LP		W	60	63	0.952	0.027
	21	D	86	120	0.717	0.041
	25	D	90	120	0.750	0.040
	29.5	D	56	79	0.709	0.051
	21	W	89	119	0.748	0.040
	25	W	93	118	0.788	0.038
	29.5	W	28	40	0.700	0.072

# **CHAPTER 4: Proportion of turtles eggs survived by Clutch**

<b>Proportion Survived</b>					
<b>Temp</b>	<b>Clutch</b>	<b>Proportion Survived</b>	<b>Temp</b>	<b>Clutch</b>	<b>Proportion Survived</b>
21	11	1.0000	29.5	11	0.8333
21	12	1.0000	29.5	12	0.8333
21	13	0.8462	29.5	13	0.8333
21	14	0.7500	29.5	14	0.6000
21	15	1.0000	29.5	15	1.0000
21	16	0.9167	29.5	16	0.8000
21	21	0.4444	29.5	21	0.3333
21	22	0.9091	29.5	22	1.0000
21	23	0.9167	29.5	23	0.0000
21	24	1.0000	29.5	24	1.0000
21	25	0.0833	29.5	25	0.0000
21	31	0.0000	29.5	31	0.1667
21	32	0.3333	29.5	32	0.6667
21	33	0.0000	29.5	33	0.0000
21	34	0.7500	29.5	34	1.0000
21	45	0.9167	29.5	45	1.0000
21	41	0.8182	29.5	41	1.0000
21	42	1.0000	29.5	42	1.0000
21	43	0.9167	29.5	43	1.0000
21	44	1.0000	29.5	44	0.8333
21	45	0.8333	29.5	45	0.8333
25	11	1.0000			
25	12	0.9000			
25	13	0.7692			
25	14	0.8333			
25	15	0.8889			
25	16	1.0000			
25	21	0.6250			
25	22	1.0000			
25	23	0.8333			
25	24	1.0000			
25	25	0.1000			
25	31	0.0833			
25	32	0.2222			
25	33	0.0714			
25	34	0.8333			
25	45	1.0000			
25	41	1.0000			
25	42	1.0000			
25	43	1.0000			
25	44	1.0000			
25	45	0.9091			

```

/* This program is used to make the main effect plots found in chapter 1*/
options linesize=96 pagesize=54 nodate;
data one;
infile 'C:\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htwt;
run;
data two; set one;
if alive in(0,1);
if inc in (1,3,4,5,6);
run;

proc freq;
tables moist*alive/nopercent nocol binomial;
tables temp*alive/nopercent nocol binomial;
tables site*alive/nopercent nocol binomial;
run;

proc tabulate data =two;
class moist temp site alive;
Table moist, (alive)*ROWPCTN;
Table temp, (alive)*ROWPCTN;
Table Site, (alive)*ROWPCTN;
Title 'Main Effect';
run;

proc summary;
Class clutch moist temp site;
output out= two N(alive)= Total Mean(alive) = meanalive std(alive)=stdalive stderr(alive)=stderro
run;
proc print;
run;

/* MAIN EFFECT PLOTS - PLOT MEANS AND STDERRORS WITH PROC GPLOT*/
Proc gplot;
plot meanalive * moist/frame hminor = 0 vminor =4 vaxis=axis1 haxis=axis2 name='GB0702';
symbol1 i=stdimjt v=none l=23 c=black;
axis1 label=(a=90 r =0 h=1.5) value=(h=1.5) offset=(4) order=(0.2 to 1.0 by 0.2);
axis2 label=(h=1.5) value=(h=1.5) offset=(4);
run;
Proc gplot data = two;
plot meanalive * temp/frame hminor = 0 vminor =4 vaxis=axis1 haxis=axis2 name='GB07021';
symbol1 i=stdimjt v=none l=23 c=black;
axis1 label=(a=90 r =0 h=1.5) value=(h=1.5) offset=(4) order=(0.2 to 1.0 by 0.2);
axis2 label=(h=1.5) value=(h=1.5) offset=(4) order = (21 to 29 by 4);
run;
Proc gplot data = two;
plot meanalive * site/frame hminor = 0 vminor =4 vaxis=axis1 haxis=axis2 name='GB07022';
symbol1 i=stdimjt v=none l=23 c=black;
axis1 label=(a=90 r =0 h=1.5) value=(h=1.5) offset=(4) order=(0.2 to 1.0 by 0.2);
axis2 label=(h=1.5) value=(h=1.5) offset=(4) ;
run;

```

```

/* This program is used to make the main effect plots found in chapter 1*/
options linesize=96 pagesize=54 nodate;
data one;
infile 'C:\Documents and Settings\Owner\Desktop\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htw;
run;
data two; set one;
if alive in(0,1);
if inc in (1,3,4,5,6);
run;

*** Print Means Breakdown for the Dependent Variables ***;
proc summary data=TWO noprint;
  class TEMP;
  VAR ALIVE;
  output out=work._aovout
    mean = _mean
    std = _std
    stderr = _stderr
    n = _n
  ;
run;
data work._aovout; set work._aovout; drop _FREQ_;
  label _mean = "Mean of ALIVE";
  label _std = "Std. Dev. of ALIVE";
  label _stderr = "Std. Error of ALIVE";
  label _n = "Number Non-missing of ALIVE";
run;
proc print data=work._aovout(drop= _type_) noobs label;
title1 'Means and Descriptive Statistics';
  id TEMP;
run;
title1;
proc delete data=work._aovout; run;
goptions reset=all device=WIN;
  title;
  footnote;
*** Means plots with standard error bars for groups *** ;
goptions ftext=SWISS ctext=BLACK htext=1 cells
  gunit=pct htitle=6;
symbol1 i=stdmtj ci=BLUE co=BLUE v=none h=1 cells width=1;
axis1 offset=(10 pct) label=(h=4) width=2 order=(21 to 29 by 4);
axis2 major=(number=5) label=(a=90 h=4) width=2;
proc gplot data=TWO;
  plot ALIVE * TEMP = 1 /
    hminor=0 vminor=0 caxis=BLACK
    description="Means plot of ALIVE by TEMP" name='means'
    haxis=axis1 vaxis=axis2;
  axis1 length=50 offset=(10 pct) label=(h=4) width=2 order=(21 to 29 by 4);
  axis2 length=50 order=(0.0 to 1.0 by 0.2);
run;
quit;
goptions reset=symbol;
axis1; axis2;

```

```

goptions reset=all device=WIN;

**** *****MOISTURE*****;
*** Print Means Breakdown for the Dependent Variables ***;
options pageno=1;
proc summary data=two noprint;
  class MOIST;
  var ALIVE;
  output out=work._aovout
    mean = _mean
    std = _std
    stderr = _stderr
    n = _n
  ;
run;
data work._aovout; set work._aovout; drop _FREQ_;
  label _mean = "Mean of ALIVE";
  label _std = "Std. Dev. of ALIVE";
  label _stderr = "Std. Error of ALIVE";
  label _n = "Number Non-missing of ALIVE";
run;
proc print data=work._aovout(drop= _type_) noobs label;
title1 'Means and Descriptive Statistics';
  id MOIST;
run;
title1;
proc delete data=work._aovout; run;
goptions reset=all device=WIN;
  title;
  footnote;
*** Means plots with standard error bars for groups *** ;
goptions ftext=SWISS ctext=BLACK htext=1 cells
  gunit=pct htitle=6;
symbol1 i=stdmtj ci=BLUE co=BLUE v=none h=1 cells width=1;
axis1 offset=(10 pct) label=(h=4) width=2;
axis2 major=(number=5) label=(a=90 h=4) width=2 order=(0.0 to 1.0 by 0.2);
proc gplot data=two ;
  plot ALIVE * MOIST = 1 /
    hminor=0 vminor=0 caxis=BLACK
    description="Means plot of ALIVE by MOIST" name='means'
    haxis=axis1 vaxis=axis2;
  axis1 length=50 offset=(10 pct)width=2;
  axis2 length=50 order=(0.0 to 1.0 by 0.2);
run;
quit;
goptions reset=symbol;
axis1; axis2;
goptions reset=all device=WIN ;

*****SITE*****;
*** Print Means Breakdown for the Dependent Variables ***;
options pageno=1;
proc summary data=TWO noprint;
  class SITE;
  var ALIVE;

```

```

output out=work._aovout
  mean = _mean
  std = _std
  stderr = _stderr
;
run;
data work._aovout; set work._aovout; drop _FREQ_;
  label _mean = 'Mean of ALIVE';
  label _std = 'Std. Dev. of ALIVE';
  label _stderr = 'Std. Error of ALIVE';
run;
proc print data=work._aovout(drop= _type_) noobs label;
  title1 'Means and Descriptive Statistics';
  id SITE;
run;
title1;
proc delete data=work._aovout; run;
goptions reset=all device=WIN;
  title;
  footnote;
*** Means plots with standard error bars for groups *** ;
goptions ftext=SWISS ctext=BLACK htext=1 cells
  gunit=pct htitle=6;
symbol1 i=std1mtj ci=BLUE co=BLUE v=none h=1 cells width=1;
axis1 offset=(10 pct) label=(h=4) width=2;
axis2 major=(number=5) label=(a=90 h=4) width=2;
proc gplot data=TWO ;
  plot ALIVE * SITE = 1 /
    hminor=0 vminor=0 caxis=BLACK
    description='Means plot of ALIVE by SITE' name='means'
    haxis=axis1 vaxis=axis2;
  axis1 length=50 offset=(5 pct) width=2 ;
  axis2 length=50 order=(0.0 to 1.0 by 0.2);
run;
quit;
goptions reset=symbol;
axis1; axis2;
goptions reset=all device=WIN ;

```

```

/* This program is used to make the interaction plots found in chapter 1*/
options linesize=96 pagesize=54 nodate;
data one;
infile 'C:\Documents and Settings\Owner\Desktop\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htw;
run;

data two; set one;
if alive in(0,1);
run;

proc summary;
Class moist temp site;
output out= two N(alive)= Total Mean(alive) = meanalive std(alive)=stdalive stderr(alive)=stderro
run;

/*Interaction Plots*/

data mts mt ms ts cs;
set two;
If _type_='110'B then output mt;
If _type_='101'B then output ms;
If _type_='011'B then output ts;
If _type_='111'B then output mts;
symbol1 v=none l=1 i=std1tmj w=2 c=black;
symbol2 v=none l=20 i=std1tmj w=2 c=red;
symbol3 v=none l=10 i=std1tmj w=2 c=blue;
symbol4 v=none l=15 i=std1tmj w=2 c=purple;

proc gplot data = mt;
plot meanalive*temp=moist/ frame vaxis=axis1 vminor=1 haxis=axis2 legend=legend1
name='GB0712';
legend1 label=(h=1.6 'Mean Alive') value =(h=1.5) shape=line(4);
axis1 label=(a=90 r=0 h=1.8) value=(h=1.5) order=(0.2 to 1.0 by 0.2)offset=(2);
axis2 label=(a=90 r=0 h=1.8) value=(h=1.5) order=(21 to 29 by 4) offset=(3) minor = none;
run;

proc gplot data = ms;
plot meanalive*site =moist/ vaxis=axis1 vminor=1 haxis=axis2 frame legend=legend1
name='GB0712';
legend1 label=(h=1.6 'Mean Alive') value =(h=1.5) shape=line(4);
axis1 label =(a=90 r=0 h=1.8) value=(h=1.5)order=(0.2 to 1.0 by 0.2)offset=(2);
axis2 label =(a=90 r=0 h=1.8) value=(h=1.5)offset=(3) minor = none;
run;

Proc gplot data = ts;
plot meanalive*temp =site/ vaxis=axis1 vminor=1 haxis=axis2 frame legend=legend1
name='GB0712';
legend1 label=(h=1.6 'Mean Alive') value =(h=1.5) shape=line(4);
axis1 label =(a=90 r=0 h=1.8) value=(h=2.0) order=(0.2 to 1.0 by 0.2)offset=(2);
axis2 label =(a=90 r=0 h=1.8) value=(h=1.7) order=(21 to 29 by 4) offset=(3) minor = none;
run;

```

```

/* This program is used to make the interaction plots found in chapter 1*/
options linesize=96 pagesize=54 nodate;
data one;
infile 'C:\Documents and Settings\Owner\Desktop\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htw;
run;
data two; set one;
if alive in(0,1);
if inc in (1,3,4,5,6);
run;

*****MOIST * SITE*****;
title1 'SITE * MOIST';
proc glm data= TWO;
  class MOIST SITE;
  model ALIVE = MOIST SITE MOIST*SITE / SS3 solution;
  ** Create output data set for plots **;
  output OUT=work._plotout p=_pred r=_resid student=_stres rstudent=_rstres
    dffits=_dffits h=_h covratio=_covr l95=_l95 u95=_u95 l95m=_l95m
    u95m=_u95m;
run; quit;
*** Print Means Breakdown for the Quantitative Variables ***;
proc summary data= TWO noprint;
  class MOIST SITE;
  var ALIVE;
  output out=work._aovout
    mean = _mean
    std = _std
    stderr = _stderr
    n = _n
    SUM = SUM
;
run;
data work._aovout; set work._aovout; drop _FREQ_;
length _effect $ 200;
_nway = length(compress(put(_type_,binary16.),'0'));
if _nway <= 2;
label _mean = "Mean of ALIVE";
label _std = "Std. Dev. of ALIVE";
label _stderr = "Std. Error of ALIVE";
label _n = "Number Non-missing of ALIVE";
if MOIST ^= '' then _effect = trimn(_effect) || "MOIST";
if SITE ^= '' then _effect = trimn(_effect) || "SITE";
if indexc(_effect,'') then _effect = substr(_effect,2);
if _effect = '' then _effect = 'Overall';
label _effect = 'Effect';
run;
proc sort data=work._aovout;
  by _nway _effect;
run;
proc print data=work._aovout(drop=_nway _type_) noobs label;
title1 'Breakdown of Means and Other Descriptive Statistics';
  id MOIST SITE;
  by _effect notsorted;

```

```

run;
title1;
proc delete data=work._aovout; run;
goptions reset=all device=WIN;
  title;
  footnote;
*** Plots ***;
goptions ftext=SWISS ctext=BLACK htext=1 cells
  gunit=pct htitle=8;
axis1 major=(number=5) label=(a=90 h=4) width=2;
axis2 offset=(10 pct) label=(h=4) width=2;
axis1 length=50;
axis2 length=50 ;
proc gplot data=work._plotout ;
  where ALIVE is not missing and MOIST is not missing and SITE is not
  missing;
** Two-way Plots **;
symbol1 i=std1mtj v=none color=cx008080 height=1 cells width=1;
symbol2 i=std1mtj v=none color=cxd08c16 height=1 cells width=1;
symbol3 i=std1mtj v=none color=cx0d0369 height=1 cells width=1;
symbol4 i=std1mtj v=none color=cx5b768d height=1 cells width=1;
symbol5 i=std1mtj v=none color=cxFF8283 height=1 cells width=1;
symbol6 i=std1mtj v=none color=cxFF00ff height=1 cells width=1;
symbol7 i=std1mtj v=none color=cx009998 height=1 cells width=1;
symbol8 i=std1mtj v=none color=cxa05000 height=1 cells width=1;
symbol9 i=std1mtj v=none color=cx2e734f height=1 cells width=1;
symbol10 i=std1mtj v=none color=cx000080 height=1 cells width=1;
plot ALIVE * SITE = MOIST /
  frame hminor=0 vminor=0 vaxis=axis1 haxis=axis2
  caxis=BLACK name='MEANS'
  description='Means plot of ALIVE by MOIST and SITE';
axis1 length=50 width=2 order=(0.0 to 1.0 by 0.2);
axis2 length=50 offset=(5 pct);
run;
quit;
proc delete data=work._plotout; run;
goptions reset=symbol ftext= ctext= htext=;
axis1; axis2; axis3;
goptions reset=all device=WIN;

*****TEMP * MOISTURE *****;
title1 'TEMP * MOIST';
*** Linear Models Analysis ***;

proc glm data=TWO;
  class TEMP MOIST;
  model ALIVE = MOIST TEMP MOIST*TEMP / SS3 solution;
  ** Create output data set for plots **;
  output OUT=work._plotout p=_pred r=_resid student=_stres rstudent=_rstres
    dffits=_dffits h=_h covratio=_covr l95=_l95 u95=_u95 l95m=_l95m
    u95m=_u95m;
run; quit;
*** Print Means Breakdown for the Quantitative Variables ***;
proc summary data=TWO noprint;
  class TEMP MOIST;

```

```

var ALIVE;
output out=work._aovout
  mean = _mean
  std = _std
  stderr = _stderr
  n = _n
SUM=SUM
;
run;
data work._aovout; set work._aovout; drop _FREQ_;
  length _effect $ 200;
  _nway = length(compress(put(_type_,binary16.),'0'));
  if _nway <= 2;
  label _mean = "Mean of ALIVE";
  label _std = "Std. Dev. of ALIVE";
  label _stderr = "Std. Error of ALIVE";
  label _n = "Number Non-missing of ALIVE";
  if TEMP ^= . then _effect = trimn(_effect) || "**TEMP";
  if MOIST ^= '' then _effect = trimn(_effect) || "**MOIST";
  if indexc(_effect,'*') then _effect = substr(_effect,2);
  if _effect = '' then _effect = 'Overall';
  label _effect = 'Effect';
run;
proc sort data=work._aovout;
  by _nway _effect;
run;
proc print data=work._aovout(drop=_nway _type_) noobs label;
title2 'Breakdown of Means and Other Descriptive Statistics';
  id TEMP MOIST;
  by _effect notsorted;
run;
title2;
proc delete data=work._aovout; run;
options reset=all device=WIN;
  title;
  footnote;
*** Plots ***;
options ftext=SWISS ctext=BLACK htext=1 cells
  gunit=pct htitle=6;
axis1 major=(number=5) label=(a=90 h=4) width=2;
axis2 offset=(10 pct) label=(h=4) width=2 ORDER = (21 TO 29 BY 4);
axis1 length=50;
axis2 length=50;
proc gplot data=work._plotout ;
  where ALIVE is not missing and TEMP is not missing and MOIST is not
  missing;
** Two-way Plots **;
symbol1 i=std1mtj v=none color=cx008080 height=1 cells width=1;
symbol2 i=std1mtj v=none color=cxd08c16 height=1 cells width=1;
symbol3 i=std1mtj v=none color=cxc0d0369 height=1 cells width=1;
symbol4 i=std1mtj v=none color=cx5b768d height=1 cells width=1;
symbol5 i=std1mtj v=none color=cxFF8283 height=1 cells width=1;
symbol6 i=std1mtj v=none color=cxFF00ff height=1 cells width=1;
symbol7 i=std1mtj v=none color=cx009998 height=1 cells width=1;
symbol8 i=std1mtj v=none color=cxa05000 height=1 cells width=1;

```

```

symbol9 i=std1mtj v=none color=cx2e734f height=1 cells width=1;
symbol10 i=std1mtj v=none color=cx000080 height=1 cells width=1;
plot ALIVE * TEMP = MOIST /
  frame hminor=0 vminor=0 vaxis=axis1 haxis=axis2
  caxis=BLACK name='MEANS'
  description='Means plot of ALIVE by MOIST and TEMP';
axis1 length=50 width=2 order=(0.0 to 1.0 by 0.2);
axis2 length=50 offset=(10 pct)order=(21 to 29 by 4);
run;
quit;

```

\*\*\*\*\*TEMP BY SITE\*\*\*\*\*

```

title1 "TEMP * SITE";
*** Linear Models Analysis ***;
proc glm data=TWO;
  class TEMP SITE;
  model ALIVE = SITE TEMP SITE*TEMP / SS3 solution;
** Create output data set for plots **;
  output OUT=work._plotout p=_pred r=_resid student=_stres rstudent=_rstres
    dffits=_dffits h=_h covratio=_covr l95=_l95 u95=_u95 l95m=_l95m
    u95m=_u95m;
run; quit;
*** Print Means Breakdown for the Quantitative Variables ***;
proc summary data=TWO noprint;
  class TEMP SITE;
  var ALIVE;
  output out=work._aovout
    mean = _mean
    std = _std
    stderr = _stderr
    n = _n
  SUM = SUM
  ;
run;
data work._aovout; set work._aovout; drop _FREQ_;
  length _effect $ 200;
  _nway = length(compress(put(_type_,binary16.),'0'));
  if _nway <= 2;
  label _mean = "Mean of ALIVE";
  label _std = "Std. Dev. of ALIVE";
  label _stderr = "Std. Error of ALIVE";
  label _n = "Number Non-missing of ALIVE";
  if TEMP ^= . then _effect = trimn(_effect) || "**TEMP";
  if SITE ^= '' then _effect = trimn(_effect) || "**SITE";
  if indexc(_effect,'*') then _effect = substr(_effect,2);
  if _effect = '' then _effect = 'Overall';
  label _effect = 'Effect';
run;
proc sort data=work._aovout;
  by _nway _effect;
run;
proc print data=work._aovout(drop=_nway _type_) noobs label;
title2 'Breakdown of Means and Other Descriptive Statistics';
  id TEMP SITE;

```

```

by _effect notsorted;
run;
title2;
proc delete data=work._aovout; run;
goptions reset=all device=WIN;
title;
footnote;

*** Plots ***;
goptions ftext=SWISS ctext=BLACK htext=1 cells
        gunit=pct htitle=6;
axis1 major=(number=5) label=(a=90 h=4) width=2;
axis2 offset=(10 pct) label=(h=4) width=2 ORDER = (21 TO 29 BY 4);
axis1 length=50;
axis2 length=50;
proc gplot data=work._plotout ;
    where ALIVE is not missing and TEMP is not missing and SITE is not missing;
** Two-way Plots **;
symbol1 i=std1mtj v=none color=cx008080 height=1 cells width=1;
symbol2 i=std1mtj v=none color=cxd08c16 height=1 cells width=1;
symbol3 i=std1mtj v=none color=cxcd0369 height=1 cells width=1;
symbol4 i=std1mtj v=none color=cx5b768d height=1 cells width=1;
symbol5 i=std1mtj v=none color=cxFF8283 height=1 cells width=1;
symbol6 i=std1mtj v=none color=cxff00ff height=1 cells width=1;
symbol7 i=std1mtj v=none color=cx009998 height=1 cells width=1;
symbol8 i=std1mtj v=none color=cxa05000 height=1 cells width=1;
symbol9 i=std1mtj v=none color=cx2e734f height=1 cells width=1;
symbol10 i=std1mtj v=none color=cx000080 height=1 cells width=1;
plot ALIVE * TEMP= SITE /
    frame hminor=0 vminor=0 vaxis=axis1 haxis=axis2
    caxis=BLACK name='MEANS'
    description='Means plot of ALIVE by SITE and TEMP';
axis1 length=50 width=2 order=(0.0 to 1.0 by 0.2);
axis2 length=50 offset=(5 pct) order=(21 to 29 by 4);
run;
quit;
proc delete data=work._plotout; run;
goptions reset=symbol ftext= ctext= htext=;
axis1; axis2; axis3;
goptions reset=all device=WIN;

```

```

/*This program was used to fit the clutch to clutch variation plot in chapter 1*/
options linesize=96 pagesize=54;
data one;
infile 'C:\turtles\turtles datafull2.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htwrt;
run;
data two; set one;
if alive in(0,1);
if inc in (1,3,4,5,6);

run;
proc tabulate;
class temp clutch alive;
tables (temp*clutch ALL),alive all;
tables (temp*clutch ALL)*rowpctn,alive all;
proc summary;
Class temp clutch;
output out=two N(alive)=Total Mean(alive) = meanalive std(alive)=stdalive stderr(alive)=stderrora
run;
proc print;
goptions reset =all;
Proc gplot;
plot meanalive*clutch=temp/ frame hminor = 2 vminor =4 vaxis=axis1 haxis=axis2;
symbol1 I=stdmt v=none l=23 c=white ;
symbol2 I=stdmt v=diamond l=23 c=green ;
symbol3 I=stdmt v=square l=23 c=red ;
symbol4 I=stdmt v=star l=30 c=blue ;
axis1 label=(a=90 r =0 h=1.5) value=(h=1.5) offset=(4) order=(0.0 to 1.0 by 0.2);
axis2 label=(h=1.5) value=(h=1.5) offset=(4);
Title 'Proportion Alive by Clutch and Temperature';
run;

/*Proc gplot;
plot meanalive*clutch/ frame hminor = 2 vminor =4 vaxis=axis1 haxis=axis2;
symbol1 I=stdmt v=none l=23 c=black ;
symbol2 I=stdmt v=diamond l=23 c=green ;
symbol3 I=stdmt v=square l=23 c=red ;
symbol4 I=stdmt v=star l=30 c=blue ;
axis1 label=(a=90 r =0 h=1.5) value=(h=1.5) offset=(4) order=(0.0 to 1.0 by 0.2);
axis2 label=(h=1.5) value=(h=1.5) offset=(4);
Title 'Proportion Alive by Clutch and Temperature';
run;

```

```

/* This program is used to analyze the whole plot in chapter 2*/
options linesize=96 pagesize=54;
data one;
infile 'C:\Documents and Settings\Owner\Desktop\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htwt;
run;
data two; set one;
if alive in(0,1);
if inc in(1,3,4,5,6);
run;
proc freq;
tables inc*alive/nopercent nocol binomial;
tables temp*alive/nopercent nocol binomial;
run;
proc freq;
tables moist*alive/nopercent nocol binomial;
tables site*alive/nopercent nocol binomial;
run;
proc summary;
class temp inc;
output out= two N(alive)= Total Mean(alive) = meanalive std(alive)=stdalive stderr(alive)=stderrora
Title ' Temperature within Incubator';
run;
proc print;
run;

Data two; set two;
input inc n y temp;
cards;
1 120 84 29
1 120 92 25
2 120 91 25
1 120 90 21
2 120 85 21
;
proc logistic;
class temp;
model y/n = temp;
run;
ods html;
ods graphics on;
proc genmod;
Class temp;
model y/n = temp/link = logit error = binomial type3 r;
output out=new resdev = residuals p=predicted ;
run;
proc plot;
plot residuals * predicted;
run;
ods graphics off;
ods html close;
proc print;
run;
/*proc tabulate;

```

```

class temp inc alive;
tables (temp*inc ALL), alive all;
tables (temp*inc ALL)*ROWPCTN, alive all;
run;*/

```

```

/* This program uses the glimmix macro to fix a generalized linear mixed model used in
chapter 3*/
options linesize=96 pagesize=54 nodate;
data one;
infile 'C:\Documents and Settings\Owner\Desktop\turtles\turtles datafull.txt';
input obs Site$ clutch$ eggid inc tray temp moist$ fert
alive stage exmth exday htmnth htday htwrt;
run;

data two; set one;
if alive in(0,1);
if inc in (1,3,4,5,6);
run;

Proc summary;
class temp inc moist tray;
output out= two N(alive)= Total sum(alive) = sum Mean(alive) = meanalive std(alive)=stdalive stderr
run;

data three;
input temp inc moist$ tray N Y;
cards;
21 1 D 1 40 30
21 1 D 2 40 29
21 1 W 1 40 31
21 2 D 1 40 27
21 2 W 1 39 26
21 2 W 2 40 32
25 1 D 1 40 33
25 1 D 2 40 30
25 1 W 1 39 29
25 2 D 1 40 27
25 2 W 1 39 29
25 2 W 2 40 35
29 1 D 1 39 26
29 1 D 2 40 30
29 1 W 1 40 28
;
run;

ods html;
ods graphics on;

%include 'C:\Documents and Settings\Owner\Desktop\Turtles\glimmix macro.sas';
%glimmix(data=three,
procopt=nobound,
stmts=%str(
class moist inc temp tray;
model Y/N = temp moist temp*moist/solution residual;
random inc(temp)/solution;
random moist*inc(temp)/solution;

estimate '21/D/ BLUP BROAD' intercept 1 temp 1 0 0 moist 1 0 temp*moist 1 0 0 0 0 0 / cl e;

estimate '21/D/ BLUP Narrow' intercept 2 temp 2 0 0 moist 2 0 temp*moist 2 0 0 0 0 0

```

```

|inc(temp) 1 1 0 0 0 moist*inc(temp) 1 1 0 0 0 0 0 0 0 / divisor = 2 e cl;

estimate '25/D/ BLUP BROAD' intercept 1 temp 0 1 0 moist 1 0 temp*moist 0 1 0 0 0 0 /e cl;

estimate '25/D/ BLUP Narrow' intercept 2 temp 0 2 0 moist 2 0 temp*moist 0 2 0 0 0 0
|inc(temp) 0 0 1 1 0 moist*inc(temp) 0 0 1 1 0 0 0 0 0 0 / divisor = 2 e cl;

),
error=binomial,
link=logit,out=setp,maxit=50
);
run;
ods graphics off;
ods html close;
proc print;
run;

```

```

/*This program again uses the glimmix macro to fit a generalized linear mixed model
as well it calculates the BLUPs all can be seen in chapter 4. This program also assumes
the factor site is fixed*/

```

```

options linesize=92 pagesize=53;

```

```

data three;

```

```

input temp inc tray moist$ site$ N Y;

```

```

cards;

```

21	1	1	D	AP	11	10
21	1	1	D	CM	9	6
21	1	1	D	CP	10	4
21	1	1	D	LP	10	10
21	1	1	W	AP	13	12
21	1	1	W	CM	8	5
21	1	1	W	CP	10	5
21	1	1	W	LP	9	9
21	1	2	D	AP	11	10
21	1	2	D	CM	10	6
21	1	2	D	CP	9	4
21	1	2	D	LP	10	9
21	2	1	W	AP	12	11
21	2	1	W	CM	9	6
21	2	1	W	CP	8	1
21	2	1	W	LP	10	8
21	2	1	D	AP	11	10
21	2	1	D	CM	9	6
21	2	1	D	CP	10	3
21	2	1	D	LP	10	8
21	2	2	W	AP	12	11
21	2	2	W	CM	8	6
21	2	2	W	CP	11	6
21	2	2	W	LP	9	9
25	1	1	D	AP	12	11
25	1	1	D	CM	7	6
25	1	1	D	CP	11	6
25	1	1	D	LP	10	10
25	1	1	W	AP	12	10
25	1	1	W	CM	8	6
25	1	1	W	CP	11	5
25	1	1	W	LP	8	8
25	1	2	D	AP	12	11
25	1	2	D	CM	9	5
25	1	2	D	CP	10	5
25	1	2	D	LP	9	9
25	2	1	W	AP	10	10
25	2	1	W	CM	8	6
25	2	1	W	CP	11	3
25	2	1	W	LP	10	10
25	2	1	D	AP	11	9
25	2	1	D	CM	9	5
25	2	1	D	CP	10	4
25	2	1	D	LP	10	9
25	2	2	W	AP	12	11
25	2	2	W	CM	10	9
25	2	2	W	CP	9	6
25	2	2	W	LP	9	9

29.5	1	1	D	AP	11	8
29.5	1	1	D	CM	8	3
29.5	1	1	D	CP	10	6
29.5	1	1	D	LP	10	9
29.5	1	1	W	AP	11	10
29.5	1	1	W	CM	10	4
29.5	1	1	W	CP	11	7
29.5	1	1	W	LP	8	7
29.5	1	2	D	AP	12	10
29.5	1	2	D	CM	8	5
29.5	1	2	D	CP	10	5
29.5	1	2	D	LP	10	10

```

;

```

```

run;

```

```

ods html;

```

```

ods graphics on;

```

```

%include 'C:\Turtles\glimmix macro.sas';

```

```

%glimmix(data=three,

```

```

    procopt=nobound,

```

```

    stmts=%str(

```

```

        class tray site moist inc temp ;

```

```

        model Y/N = temp moist temp*moist site site*temp site*moist/residual;

```

```

        random inc(temp)/s;

```

```

        random tray(moist*temp*inc)/solution;

```

```

    title 'Blup Inference when Sites is fixed';

```

```

    estimate '21/D/AP BLUP BROAD' intercept 1 temp 1 0 0 moist 1 0 temp*moist 1 0 0 0 0
        site 1 0 0 0 site*temp 1 0 0 0 0 0 0 0 0 0 site*moist 1 0 0 0 0

```

```

    estimate '21/D/AP/ BLUP Narrow' intercept 6 temp 6 0 0 moist 6 0 temp*moist 6 0 0 0 0
    site 6 0 0 0 temp*site 6 0 0 0 0 0 0 0 0 0 site*moist 6 0 0 0 0 0 0
    |inc(temp) 3 3 0 0 0 tray(moist*temp*inc) 2 2 0 0 0 2 0 0 0 / divisor = 6 e;

```

```

    estimate '25/D/AP BLUP BROAD' intercept 1 temp 0 1 0 moist 1 0 temp*moist 0 1 0 0 0
        site 1 0 0 0 site*temp 0 1 0 0 0 0 0 0 0 0 site*moist 1 0 0 0 0

```

```

    estimate '25/D/AP/ BLUP Narrow' intercept 6 temp 0 6 0 moist 6 0 temp*moist 0 6 0 0 0
    site 6 0 0 0 temp*site 0 6 0 0 0 0 0 0 0 0 site*moist 6 0 0 0 0 0 0
    |inc(temp) 0 0 3 3 0 tray(moist*temp*inc) 0 0 2 2 0 0 2 0 0 / divisor = 6 e;

```

```

    estimate '21/D/CM/ BLUP Narrow' intercept 6 temp 6 0 0 moist 6 0 temp*moist 6 0 0 0 0
    site 0 6 0 0 temp*site 0 0 6 0 0 0 0 0 0 0 site*moist 0 0 6 0 0 0 0
    |inc(temp) 3 3 0 0 0 tray(moist*temp*inc) 2 2 0 0 0 2 0 0 0 / divisor = 6 e;

```

```

);

```

```

error=binomial,

```

```

link=logit,out=setp,maxit=50

```

```

);

```

```

run;

```

```

proc print;

```

```

run;

```

```

ods graphics off;

```

```

ods html close;

```

```

/*This program again uses the glimmix macro to fit a generalized linear mixed model
as well it calculates the BLUPs all can be seen in chapter 4. This program also assumes
the factor site is random*/
options linesize=92 pagesize=53;
data three;
input temp inc tray moist$ site$ N Y;
cards;
21 1 1 D AP 11 10
21 1 1 D CM 9 6
21 1 1 D CP 10 4
21 1 1 D LP 10 10
21 1 1 W AP 13 12
21 1 1 W CM 8 5
21 1 1 W CP 10 5
21 1 1 W LP 9 9
21 1 2 D AP 11 10
21 1 2 D CM 10 6
21 1 2 D CP 9 4
21 1 2 D LP 10 9
21 2 1 W AP 12 11
21 2 1 W CM 9 6
21 2 1 W CP 8 1
21 2 1 W LP 10 8
21 2 1 D AP 11 10
21 2 1 D CM 9 6
21 2 1 D CP 10 3
21 2 1 D LP 10 8
21 2 2 W AP 12 11
21 2 2 W CM 8 6
21 2 2 W CP 11 6
21 2 2 W LP 9 9
25 1 1 D AP 12 11
25 1 1 D CM 7 6
25 1 1 D CP 11 6
25 1 1 D LP 10 10
25 1 1 W AP 12 10
25 1 1 W CM 8 6
25 1 1 W CP 11 5
25 1 1 W LP 8 8
25 1 2 D AP 12 11
25 1 2 D CM 9 5
25 1 2 D CP 10 5
25 1 2 D LP 9 9
25 2 1 W AP 10 10
25 2 1 W CM 8 6
25 2 1 W CP 11 3
25 2 1 W LP 10 10
25 2 1 D AP 11 9
25 2 1 D CM 9 5
25 2 1 D CP 10 4
25 2 1 D LP 10 9
25 2 2 W AP 12 11
25 2 2 W CM 10 9
25 2 2 W CP 9 6
25 2 2 W LP 9 9

```

```

29.5 1 1 D AP 11 8
29.5 1 1 D CM 8 3
29.5 1 1 D CP 10 6
29.5 1 1 D LP 10 9
29.5 1 1 W AP 11 10
29.5 1 1 W CM 10 4
29.5 1 1 W CP 11 7
29.5 1 1 W LP 8 7
29.5 1 2 D AP 12 10
29.5 1 2 D CM 8 5
29.5 1 2 D CP 10 5
29.5 1 2 D LP 10 10

;
run;
ods html;
ods graphics on;
%include 'C:\Turtles\glimmix macro.sas';
%glimmix(data=three,
procopt=nobound,
stmts=%str(
class tray site moist inc temp ;
model Y/N = temp moist temp*moist /solution residual;
random site site*temp site*moist/s;
random inc(temp)/s;
random tray(moist*temp*inc)/solution;
title ' Blup Inference when Sites is Random';
estimate '21/D/ BLUP Broad' intercept 1 temp 1 0 0 moist 1 0 temp*moist 1 0 0 0 0 / e;

estimate '21/D/AP BLUP Narrow' intercept 12 temp 12 0 0 moist 12 0 temp*moist 12 0 0 0 0
|inc(temp) 6 6 0 0 0 tray(moist*temp*inc) 4 4 0 0 0 4 0 0 0 site 12 0 0 0
temp*site 12 0 0 0 0 0 0 0 0 0 site*moist 12 0 0 0 0 0 0 / divisor = 12 e;

estimate '21/D/CM BLUP Narrow' intercept 12 temp 12 0 0 moist 12 0 temp*moist 12 0 0 0 0
|inc(temp) 6 6 0 0 0 tray(moist*temp*inc) 4 4 0 0 0 4 0 0 0 site 0 12 0 0
temp*site 0 0 0 12 0 0 0 0 0 0 site*moist 0 0 12 0 0 0 0 / divisor = 12 e;

estimate '21/D/CP BLUP Narrow' intercept 12 temp 12 0 0 moist 12 0 temp*moist 12 0 0 0 0
|inc(temp) 6 6 0 0 0 tray(moist*temp*inc) 4 4 0 0 0 4 0 0 0 site 0 0 12 0
temp*site 0 0 0 0 0 12 0 0 0 0 site*moist 0 0 0 0 12 0 0 / divisor = 12 e;

estimate '21/D/LP BLUP Narrow' intercept 12 temp 12 0 0 moist 12 0 temp*moist 12 0 0 0 0
|inc(temp) 6 6 0 0 0 tray(moist*temp*inc) 4 4 0 0 0 4 0 0 0 site 0 0 0 12
temp*site 0 0 0 0 0 0 0 0 12 0 0 site*moist 0 0 0 0 0 12 0 / divisor = 12 e;

),
error=binomial,
link=logit, out=setp,maxit=50
);
run;
proc print;
run;
ods graphics off;
ods html close;

```

## CASE STUDY #1

The effects of incubation temperature and moisture on the survival of snapping turtle (*Chelydra serpentina*) embryos

Experiment performed by Michele Bobyn  
under the supervision of Dr. R. J. Brooks

Objective: Determine the effect of  
1) temperature  
2) moisture  
3) site  
on the embryonic survival of snapping turtle  
embryos and how embryonic survival varies from  
clutch to clutch.

It is hypothesized that more extreme temperatures or very dry substrate may adversely affect survival of embryos until hatching and that some clutches or even population sites (eg. Cootes' Paradise) may be more prone to embryo death prior to hatching regardless of environmental conditions during incubation. For example, toxic load.

## Materials and Methods

Eggs from snapping turtle clutches were collected in early June, 1988, within 24 h. of oviposition from 4 Ontario nesting sites in the following areas: the North Madawaska drainage system of Algonquin Park (45 35'N, 78 30'W) (6 clutches-these eggs were individually weighed), Cootes' Paradise (43 17'N, 79 53'W) near Hamilton (5 clutches), Big Creek Marsh (42 36'N, 80 27'W) near Long Point (5 clutches) and Cranberry Marsh (?,?) near Ajax (5 clutches). The eggs were individually labelled with their oviposition site (AP, CP, LP or CM), clutch identification and egg number (#1=last laid egg) with a fine-tipped permanent felt marker, or pencil. Eggs from Cranberry Marsh were collected from female turtles injected with oxytocin to induce release of their eggs. Each clutch was arranged in a single layer in a 32.3x20.5x12.6 cm plastic shoe-box mouse cage, embedded in and lightly

covered with a 1:1.1 mixture of vermiculite:water (or damp sand in the case of the CM eggs); the cage was loosely sealed with aluminum foil to prevent desiccation on the trip back to the University of Guelph. All eggs were kept at a cool temperature (20 C) prior to being placed in incubators.

On July 6, 1988, a total of 720 eggs from 6 AP, 5 CP, 5 LP and 5 CM clutches were randomly distributed into 18 covered plastic boxes (33.4x22.6x5.9 cm: Durphy Packaging Co., Huntingdon Valley, Pa. USA) which contained 200 g of room humidity, medium-size "Terra-Lite" vermiculite. Nine boxes (or trays) were randomly assigned to each of 2 moisture contents:

1) WET: 342.86 g water added to the 200 g vermiculite in the tray (63% humidity)

2) DRY: 85.7 g water added to 200 g vermiculite (30% humidity)

Each box held 40 eggs such that 2 eggs from each clutch (wherever possible) were in each box. The eggs were randomly arranged in eight rows of five eggs each and were positioned so that half of the egg was exposed to the air within the box and the other half buried in the vermiculite. Each egg was oriented so that the white spot (indicating adherence of membranes to the interior surface of the egg shell) was uppermost, and care was taken to keep the eggs this way. Only those eggs that had distinct white spots were selected for incubation. Each box with its eggs, vermiculite and water was weighed before being put into the incubators. There were 6 incubators (Koolatron cooler-warmers, Koolatron Corp., Brantford, Ont.), 2 set to each of three temperatures (29.5 C, 25 C and 21 C). Three boxes were placed in

each incubator such that an incubator contained either two high moisture boxes (trays) and one low moisture tray (HHL), or one high moisture tray and two low moisture trays (LLH); each temperature contained a HHL and LLH group of trays. The temperature and moisture conditions were not considered to be very extreme, as they have been recorded in natural nests. The three trays were stacked vertically within the Koolatron. Below them was kept a small 15x6 cm aluminum tray full of water in order to keep humidity inside each incubator relatively constant. During the incubation of the eggs, every Monday, Wednesday and Friday, temperature readings from a stick mercury thermometer inside each Koolatron were recorded. In addition, the top box of eggs was switched to the bottom position and the entire stack of three boxes was rotated 180 degrees so that the eggs were exposed to the range of temperature variation within that Koolatron.

All 6 Koolatron units were placed in a row on a table inside a temperature-controlled room (19-20 C); room humidity was not controlled. To monitor the moisture level in the boxes, all the eggs from a given box were removed to a temporary tray, the box with its moist vermiculite was weighed, and sufficient water was added to return the box assembly to its original weight. As the eggs were removed, they were weighed to 0.1 g. They were then replaced in their original positions. This procedure was followed weekly throughout the course of incubation. Except during the rotation of the trays and the weekly weighings, the eggs were not disturbed. If eggs were dropped or otherwise accidentally disturbed somewhat more severely than usual, it was noted (usually this did not seem to result in any adverse effects in hatching, deformities etc. except in the case of some very large

eggs in the WET substrate which burst on contact while being weighed).

As soon as the first signs of hatching were apparent (i.e. pipping of eggs or complete emergence of a hatchling from its eggshell), all the eggs from that Koolatron were removed and placed individually in 100 ml (4.4x7.1 cm) glass jars which contained 7.0 g of vermiculite and either 12.0 g of water (WET) or 3.0 g of water (DRY). The lids of the jars were labelled with the egg's identity, screwed on and loosened by 1/4 turn to allow for respiration. These jars were stored in a separate room set at 20 C. Upon hatching, the turtle was removed from the jar, rinsed of adhering vermiculite and egg membranes, patted dry on paper towels and weighed to 0.1 g. Measurements of carapace length, carapace height and plastron length were also taken, and the turtle was returned to a clean jar half full of water. Notes of any obvious morphological deformities were taken. Each hatchling was then individually tagged with small loops of .020 gauge steel fishing wire through one or more posterior marginal scutes of the carapace. After being tagged, the turtles were placed in one of ten 20 gallon aquaria in 5 cm of tap water, with approximately 40 turtles per tank. They were then moved to a temperature-controlled room (25 C) with a 13 h:11 h light/dark cycle. They are presently being fed weekly with a mixture of chopped chicken or pork heart, bone meal powder, crushed reptile food pellets and bloodworms) in the amount of .25-.5 g/turtle (essentially ad lib) while they acclimate to the food, light and temperature conditions. The turtles will soon be grouped, as they were in the incubators, into one of 5 tanks.

There will only be 5, not 6, tanks used because, less than a week (July 12) after the start of the artificial incubation, one Koolatron experienced a mysterious mechanical failure resulting in the loss of 120 eggs. A total of 495 eggs hatched and 416 hatchlings survive at present. During incubation, those eggs that appeared dead, crumpled or infected with fungus were removed from the boxes, opened and examined. Other viable eggs were not disturbed. If present, the embryo was identified as having died in one of the following developmental stages:

0	egg was infertile
5	stages 0-10; some development has occurred
11-25	various stages
26	died at hatching, while hatching
27	survived hatching

If the embryo died at a late developmental stage (>22), its sex was determined if possible. The plastron and viscera were dissected away to expose the gonads and oviducts (if present), and a preliminary sex determination was made based on gross morphology. Each dead hatchling was individually labelled with a vinyl loop label (C. Frensch Ltd., Grimsby, Ont.) slipped over its head, and preserved in 70% ethanol. A second "blind" sex determination will also be made. Eggs that failed to hatch and turtles that died after hatching were similarly treated.

SITE	CLUTCH ID	NO. EGGS IN CLUTCH (initially)
AP	B7	40
	K7	31
	L11	39
	N7	36
	S10	34
CP	X6	32
	21	36
	28	29
	69	36
	80	41
LP	301	40
	100	33

	176	34
	179	35
	184	35
	190	35
CM	1	24
	9	32
	132	34
	134	35
	137	29

Yntema's embryonic snapping turtle stages at 20 C incub. temp.

STAGE	# DAYS
0	laying
1	1
2	2
3	3
4	4
5	5
6	7
7	9
8	12
9	16
10	20
11	25
12	30
13	35
14	42
15	49
16	56
17	63
18	70
19	77
20	84
21	91
22	98
23	105
24	119
25	133
26	140
(27	survived past hatching)

\*Total incubation time decreases as incubation temperature increases, therefore time between stages is compressed.

Objective: Determine the effect of

- 1)temperature
- 2)moisture
- 3)site

on the embryonic survival of snapping turtle embryos and how embryonic survival varies from clutch to clutch.



therefore, developed for @ 1 month at cool temp. and 6 days at hot temp. which would correspond roughly to stages 13-15. Taking into account the actual days each clutch was laid on (June 6-22), the stages would correspond.

If the stage is 5 this means that the embryo died in early development when the precise stage is difficult to ascertain. Presumably the one stage 5 embryo from the incubator that failed died naturally before the incubator cooked the rest of the eggs. It is possible that some of the eggs in that incubator died naturally but I think, given the stages they were at, most if not all were due to the malfunction.

Thanks for the questions. I hope this makes things a bit more clear.

Michele Bobyn

\*\*\* 1. For some reason, this data line is missing:

LP 190 35 4 3 25 W 1 1 27 . . 9 4 9.0

(site, clutch, eggid, inc, tray, temp, moist, fert, alive,  
stage, exammonth, examday, hatchmonth, hatchday, hatchwt)

2. "alive" code=2 if the egg was accidentally broken or if it burst of its own accord. Obs. 210 was broken by me; obs. 268 and 389 burst in the incubator; obs. 503 looked fungoid and dead so was dissected and a live embryo found inside. The egg was not fungoid.

3. Obs. 543 - egg was fungoid and dead but was not staged or examined

4. obs. 76 and 197 died while hatching and so were not weighed as they were still attached to and had not resorbed their yolk sacs. Obs. 177 hatched but was not weighed and it died immediately after hatching. Obs. 318 was simply not weighed, by mistake...

## ***Bibliography***

1. Bain, Lee J., Engelhardt, Max. *Introduction to probability and mathematical statistics 2<sup>nd</sup> ed.*, Duxbury
2. Brown, Jason. *GLIMMIX macro for Version 8*,  
<http://ewe.3sas.cp/techsup/download/stat/glmm800.html>
3. Bobyn, Michelle. *The effects of incubation temperature and moisture on the survival of snapping turtle (Chelydra serpentina) embryos*. University of Guelph., 1988
4. Breslow, N. E. and Clayton, D.G. *Approximate Inference in Generalized Linear Mixed Models* Journal of the American Statistical Association. March 1993, Vol. 88.
5. Littell, Ramon C., Milliken, George A., Stroup, Walter W., and Wolfinger, Russell D., *SAS<sup>®</sup> System for Mixed Models*, Cary, NC: SAS Institute Inc., 1996.
6. *Map of Ontario* - <http://atlas.gc.ca>
7. McCullagh, P. and Nelder, J.A. *Generalized Linear Models*, Chapman and Hall, London. 1987
8. McCullagh, Charles E., Searle, Shayle R., *Generalized, Linear, and Mixed Models*, Wiley series in probability and statistics 2001
9. Myers, Raymond H., Montgomery, Douglas C., Vining Geoffrey G., *Generalized linear models: with applications in engineering and the sciences*, Wiley series in probability and statistics 2002..
10. Pierce, F.J., Schabenberger, Oliver., *Contemporary statistical models for the plant and social sciences*, CRC Press, 2002 . 738 pp.
11. SAS Institute Inc.-Online Documentation: <http://v9doc.sas.com/sasdoc>