In this lab we will try to clarify the topic of link function estimation in traditional generalized linear models, a topic that was basically rushed through in lecture. The underlying idea here is that we may formulate a generalized linear model in the usual manner except that, rather than specifying a completely known link function such as log(µ), we specify a link function only be contained in some family of link functions that are indexed by a parameter, λ say. The link function then becomes \( g(\mu|\lambda) \) and we may wish to estimate \( \lambda \) along with the other parameters of the model. We will illustrate such link function estimation with analysis of data from a short-term toxicity test.

**Historical Note**

To the best of my knowledge, the concept of embedding a link function into an entire family of functions was introduced by Pregibon (1980; “Goodness of link tests for generalized linear models”, Applied Statistics 29, 15-24). Pregibon used this idea primarily to develop a score test for a hypothesized link function. The idea was that, we will typically have a “hypothesized” link function (the one we’re thinking about using) and would like to assess that hypothesized link against a range of alternatives. Suppose that the hypothesized link function can be embedded in a parameterized family of link functions in the following way:

- Suppose there is a “true” link function that is a member of some parameterized family of links, say \( g(\mu|\lambda) \). For example, suppose that the parameterized family of link functions is the power family,

\[
g(\mu|\lambda) = \begin{cases} 
\mu^\lambda & \text{if } \lambda \neq 0 \\
\log(\mu) & \text{if } \lambda = 0
\end{cases}
\]
and the true (but unknown) link function belongs to this family for a particular value of $\lambda$, say $\lambda^*$.  

- Now suppose also that we are considering using a model with a link function in the above family with a particular value of the parameter $\lambda_0$; $\lambda = 1$, for example would give an identity link. We might like a test for the hypothesis that $\lambda^* = \lambda_0$, and this is the problem Pregibon addressed.

Pregibon’s solution made clever use of a first order Taylor series for the true link function expanded about the hypothesized link function, resulting in a model that could be “fitted” for one step using the hypothesized link (meaning no new software was needed, which was more important in 1980 than it is now). Pregibon also noted that this procedure was the first step of what could become an iterative solution for maximum likelihood estimation of $\lambda$ which was perhaps unfortunate, because that procedure has been used in an improper manner under the assumption that maximum likelihood estimates resulted (see Kaiser (1997) “Maximum likelihood estimation of link function parameters”, Computational Statistics and Data Analysis 24, 79-87). But, there is an easy way to do it right, which we now give.

**Maximum Likelihood Estimation**

Let $Y_1, \ldots, Y_n$ represent independent response variables from exponential dispersion family distributions with density or mass functions,

$$f(y|\theta_i, \phi) = \exp \left[ a(\phi) \{y\theta_i - b(\theta_i)\} + c(y, \phi) \right], \quad (1)$$

where $a(\phi) = \phi m_i$ for a known set of “weights” $\{m_i : i = 1, \ldots, n\}$. We have modified our standard form from class just a bit by using this function $a(\phi)$ in place of the simple $\phi$; this will be useful in considering binomial random components (in which case the $m_i$ become the binomial “sample sizes”). To complete the model, let
the systematic model component be written as,

\[ g(\mu|\lambda) = x_i^T \beta = \eta_i \]  

(2)

where \( g(\mu|\lambda) \) is some family of link functions specified up to an unknown parameter \( \lambda \), which may be either a scalar or vector. The \( i^{th} \) contribution to the log likelihood now is,

\[ L_i = a(\phi)\{y_i - b(\theta_i)\} + c(y, \phi), \]  

(3)

and the complete log likelihood is \( L = \sum L_i \) where the sum runs from \( i = 1, \ldots, n \).

Here, we suppose that \( x_i^T = (x_{i,1}, \ldots, x_{i,p}) \), \( \beta = (\beta_1, \ldots, \beta_p)^T \), and \( \lambda = (\lambda_1, \ldots, \lambda_q)^T \). Following the same type of progression we used in class for developing a Fisher Scoring algorithm to locate maximum likelihood estimates of the regression parameters \( \beta \), we have

\[ \frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \]

\[ \frac{\partial L_i}{\partial \lambda_k} = \frac{\partial L_i}{\partial \theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\lambda_k}. \]  

(4)

Note that in (4) the third right hand side terms are now partial derivatives rather than the ratio of differentials used in class for standard generalized linear models.

Now, in a way similar to what was done in class, note that

\[ \frac{\partial L_i}{\partial \theta_i} = a(\phi)\{y_i - b'(\theta_i)\} = a(\phi)(y_i - \mu_i) \]

\[ \frac{d\theta_i}{d\mu_i} = \left( \frac{d\mu_i}{d\theta_i} \right)^{-1} = \frac{1}{b''(\theta_i)} = V^{-1}(\mu_i) \]

\[ \frac{\partial \mu_i}{\partial \eta_i} = \left( \frac{\partial g(\mu_i|\lambda)}{\partial \mu_i} \right)^{-1} = \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} \]

\[ \frac{\partial \eta_i}{\partial \beta_j} = x_{i,j} \]

\[ \frac{\partial \mu_i}{\partial \lambda_k} = \left( -\frac{\partial g(\mu_i|\lambda)}{\partial \lambda_k} \right) \left( \frac{\partial g(\mu_i|\lambda)}{\partial \mu_i} \right)^{-1} = \left( -\frac{\partial \eta_i}{\partial \lambda_k} \right) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} \]  

(5)
Note: In the third and fifth lines of expression (5) we have applied an implicit function theorem as follows. First, note that for a standard generalized linear model it is easy to write $$\mu_i = g^{-1}(\eta_i)$$ where $$g^{-1}(\cdot)$$ is the inverse function of the link $$g(\cdot)$$. When $$g(\cdot)$$ is a simple function of one argument this is not difficult (e.g., if $$g(x) = \log(x)$$ then $$g^{-1}(x) = \exp(x))$$. But when the link function is parameterized as $$g(\cdot|\lambda)$$ this is often not possible. Nevertheless, it remains true that $$g(\mu_i|\lambda) - \eta_i = 0$$ and then implicit functions immediately give lines three and five.

From the expressions (4) and (5) we can now write the $$i^{th}$$ contribution to the first derivatives as,

$$\frac{\partial L_i}{\partial \beta_j} = \phi m_i(y_i - \mu_i) V^{-1}(\mu_i) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} x_{i,j}$$

$$\frac{\partial L_i}{\partial \lambda_k} = \phi m_i(y_i - \mu_i) V^{-1}(\mu_i) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} \left( -\frac{\partial \eta_i}{\partial \lambda_k} \right)$$

Now define the terms

$$w_i = m_i \left[ \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(\mu_i) \right]^{-1}$$

which allows the above expressions to be written as,

$$\frac{\partial L_i}{\partial \beta_j} = \phi (y_i - \mu_i) w_i \left( \frac{\partial \eta_i}{\partial \mu_i} \right) x_{i,j}$$

$$\frac{\partial L_i}{\partial \lambda_k} = \phi (y_i - \mu_i) w_i \left( \frac{\partial \eta_i}{\partial \mu_i} \right) \left( -\frac{\partial \eta_i}{\partial \lambda_k} \right)$$

(6)

The point here is that the contribution of individual terms to the score functions (first derivatives) have been written in the same form for derivatives with respect to the link function parameters as for the regression parameters. That is, the only difference between the first and second lines of (6) is the final term on the right hand side.

Now, following the same progression for second derivatives and taking expected values to simplify the expressions, which is laid out in some detail in the course notes.
in Section 8.3.6 (starting on page 420), we end up with the following expressions.

\[-E \left( \frac{\partial^2 L_i}{\partial \beta_j \partial \beta_l} \right) = \phi w_i x_{i,j} x_{i,l} \]

\[-E \left( \frac{\partial^2 L_i}{\partial \beta_j \partial \lambda_k} \right) = \phi w_i \left( \frac{-\partial \eta_l}{\partial \lambda_k} \right) x_{i,j} \]

\[-E \left( \frac{\partial^2 L_i}{\partial \lambda_k \partial \lambda_h} \right) = \phi w_i \left( \frac{-\partial \eta_l}{\partial \lambda_k} \right) \left( \frac{-\partial \eta_l}{\partial \lambda_h} \right) \tag{7} \]

Now, let \( \xi \equiv (\beta^T, \lambda^T)^T \) be the complete \((p + q)\) vector of systematic model component parameters. Summing over the expressions in (6) and collecting the score functions into a vector results in the gradient,

\[ \nabla L = \left( \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_p}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \lambda_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \lambda_q} \right)^T \tag{8} \]

The (negative) expected second derivatives may be collected into a matrix \( H \) as

\[ H = \begin{pmatrix} H_1 & H_2 \\ H_2^T & H_3 \end{pmatrix}, \tag{9} \]

where

\[ H_1 \text{ is } p \times p \text{ with } j^{th} \text{ element } \sum_{i=1}^{n} \phi w_i x_{i,j} x_{i,d} \]

\[ H_2 \text{ is } p \times q \text{ with } jk^{th} \text{ element } \sum_{i=1}^{n} \phi w_i \left( \frac{-\partial \eta_l}{\partial \lambda_k} \right) x_{i,j} \]

\[ H_3 \text{ is } q \times q \text{ with } k\ell^{th} \text{ element } \sum_{i=1}^{n} \phi w_i \left( \frac{-\partial \eta_l}{\partial \lambda_k} \right) \left( \frac{-\partial \eta_l}{\partial \lambda_{\ell}} \right) \tag{10} \]

A Fisher Scoring algorithm may then be defined to update a current estimate \( \xi^{(m)} \) to a new estimate \( \xi^{(m+1)} \) as,

\[ \xi^{(m+1)} = \xi^{(m)} + \left( H^{-1} \nabla L \right) \big|_{\xi=\xi^{(m)}}. \tag{11} \]

This is really all we need to locate maximum likelihood estimates of the elements of \( \xi \) but, as with standard generalized linear models (see Section 8.3.6 in course notes)
we can perform some further manipulations to arrive at the form of an iteratively weighted least squares algorithm as follows.

Let \( X_A \) denote a matrix formed by augmenting the usual \( X \) matrix with \( q \) additional columns having elements given by the derivatives

\[
-\frac{\partial \eta_i}{\partial \lambda_k}; \quad k = 1, \ldots, q
\]

Then \( X_A \) is an \( n \times (p + q) \) matrix with \( i \)th row

\[
x_{A,i}^T = (x_{1,i}, \ldots, x_{p,i}, -\frac{\partial \eta_i}{\partial \lambda_1}, \ldots, -\frac{\partial \eta_i}{\partial \lambda_q})
\]

Let \( W \) be an \( n \times n \) diagonal matrix with elements \( w_i \) as given immediately prior to expression (6), and let \( z^T = (z_1, \ldots, z_n) \) where

\[
z_i = (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}.
\]

Then we have that

\[
\nabla L = \phi X_A^T W z \quad \text{and} \quad H = \phi X_A^T W X_A,
\]

and the Fisher Scoring algorithm on expression (11) then becomes

\[
\xi^{(m+1)} = \xi^{(m)} + \left[ (X_A^T W X_A)^{-1} X_A^T W z \right] |_{\xi = \xi^{(m)}}
\]

\[
= \left[ (X_A^T W X_A)^{-1} (X_A^T W \xi + X_A^T W z) \right] |_{\xi = \xi^{(m)}}
\]

\[
= \left[ (X_A^T W X_A)^{-1} X_A^T W z^* \right] |_{\xi = \xi^{(m)}}, \quad (13)
\]

where, \( z^* = (z_1^*, \ldots, z_n^*)^T \) with

\[
z_i^* = x_{A,i}^T \xi + z_i
\]

The last line of (13) is in the form of an iteratively re-weighted least squares algorithm. Note here that the matrix \( X_A \) does not remain fixed in this algorithm as
it needs to be evaluated at the current estimate $\xi^{(m)}$ at each iteration.

**Useful Families of Link Functions**

If we would like to estimate parameters of link functions we need useful families of such functions, and developing these is not a trivial task. Consider, for example, the power family given earlier in this lab,

$$g(\mu|\lambda) = \begin{cases} 
\mu^\lambda & \lambda \neq 0 \\
\log(\mu) & \lambda = 0 
\end{cases} \quad (14)$$

This is a fine way to write a family of functions if our use is to select a power for a fixed link, but it is not so useful for estimation (of $\lambda$) if our desire is to separate a log link from some other power. For this, we might consider a family of link functions given by Pregibon (1980) as,

$$g(\mu|\lambda) = \frac{1}{\lambda_2} \left[ (\mu + \lambda_1)^{\lambda_2} - 1 \right]. \quad (15)$$

This family includes, for example, the identity link which results from taking $\lambda_1 = \lambda_2 = 1$. It also includes the log link if we take $\lambda_1 = 0$ and let $\lambda_2 \to 0$,

$$\lim_{\lambda_2 \to 0} g(\mu|\lambda_1 = 0, \lambda_2) = \lim_{\lambda_2 \to 0} \frac{\mu^{\lambda_2} - 1}{\lambda_2} = \log(\mu).$$

Note that the difference with the power family (14) is that in (14) the link was defined as log for $\lambda = 0$ while in (15) we get the log link as the value of $\lambda_2$ goes to zero. This makes a difference if our objective is to estimate $\lambda$. What about other powers? Consider using $\lambda_1 = 0$ and $\lambda_2 = 2$ in (15). This gives

$$g(\mu|\lambda) = \frac{\mu^2 - 1}{2}.$$  

Now, suppose that the linear predictor is $\eta_i = \beta_0 + \beta_1 x_i$. Then we would have the systematic model component,

$$g(\mu_i|\lambda) = \beta_0 + \beta_1 x_i \Rightarrow \frac{\mu_i^2 - 1}{2} = \beta_0 + \beta_1 x_i \Rightarrow \mu_i^2 = (2 \beta_0 + 1) + 2 \beta_1 x_i.$$
or,
\[ \mu_i^2 = \gamma_0 + \gamma_1 x_i \]
so that our model is one with an ordinary squared link function.

Another useful family of link functions that we will use in analysis of a short-term toxicity test is,
\[
g(\mu|\lambda) = \log \left[ \frac{1}{(1-\mu)^\lambda} - 1 \right] - \log(\lambda).
\]
This family includes the logit link for \( \lambda = 1 \),
\[
g(\mu|\lambda = 1) = \log \left[ \frac{1}{(1-\mu)} - 1 \right] = \log \left( \frac{\mu}{1-\mu} \right),
\]
and the complementary log-log link as \( \lambda \to 0 \),
\[
\lim_{\lambda \to 0} g(\mu|\lambda) = \lim_{\lambda \to 0} \log \left[ \frac{1}{(1-\mu)^\lambda} - 1 \right] - \log(\lambda)
\]
\[
= \log \left[ \lim_{\lambda \to 0} \frac{1}{\lambda} \left\{ \frac{1}{(1-\mu)^\lambda} - 1 \right\} \right]
\]
\[
= \log \left[ \lim_{\lambda \to 0} \frac{1 - (1-\mu)^\lambda}{\lambda(1-\mu)^\lambda} \right]
\]
\[
= \log \left[ \lim_{\lambda \to 0} \frac{-\log(1-\mu)}{1 + \lambda \log(1-\mu)} \right]
\]
\[
= \log \left[ -\log(1-\mu) \right].
\]

**Analysis of Short-Term Toxicity Test Data**
The design of a short-term toxicity test is quite simple. We have \( k \) concentrations of some potentially toxic substance and we expose groups of \( n_1, n_2, \ldots, n_k \) organisms to these concentrations for a fixed period of time. At the end of that time the number
of organisms in each group that have “responded” (usually died) is recorded. Such data are often called quantal response data.

The theoretical basis for the analysis of quantal response data is somewhat more complex than the experimental design. The fundamental elements of this theory are as follows.

1. It is supposed that for each individual organism there is a concentration of the toxicant, \( x \) say, such that the organism will respond for any concentration greater than or equal to \( x \) and the organism will not respond for any concentration less than \( x \). This value is called the \textit{tolerance} of the organism. That is, if \( R_j \) denotes the response of organism \( j \), \( x_j \) its tolerance, and \( d \) the concentration to which it is exposed,

\[
Pr(R_j = 1 | d < x_j) = 0 \quad Pr(R_j = 1 | d \geq x_j) = 1
\]

2. It is also supposed that, in the population of organisms, the tolerances \( x_j \) follow some distribution that is a location-scale family with distribution function \( G \), mean \( \mu_x \) and variance \( \sigma_x^2 \). Then the standardized tolerances are such that,

\[
\bar{x}_j = \frac{x_j - \mu_x}{\sigma_x} \sim iidG(0, 1)
\]

3. Assume that at a given experimental concentrations (dose) \( d_i; i = 1, \ldots, k \), there is a certain probability \( p_i; i = 1, \ldots, k \) that the dose will exceed the tolerance of a randomly chosen organism so that, if \( Y_i \) is defined as the number of responses out of \( n_i \) organisms at dose \( i \), the probability mass function of \( Y_i \) is,

\[
f(y_i|p_i) = \frac{n_i!}{y_i!(n_i - y_i)!} p_i^{y_i} (1 - p_i)^{n_i - y_i}; \quad y_i = 0, 1, \ldots, n_i \quad (16)
\]

4. From the above we have that, at a given dose \( d_i \),

\[
p_i = \int_{-\infty}^{\delta_i} dG, \quad \text{where} \quad \delta_i = \frac{d_i - \mu_x}{\sigma_x}.
\]
That is, $\delta_i$ is the $p(100)\%$--tile of $G(0, 1)$, the distribution of standardized tolerances.

5. The objective is, given fixed $d_1, \ldots, d_k$, fixed $n_1, \ldots, n_k$ and observed $y_1, \ldots, y_k$, estimate $\mu_x$ and $\sigma^2_x$, the parameters of the tolerance distribution $G(\mu_x, \sigma^2_x)$. In particular, if the response is mortality and $G$ is assumed or chosen to be symmetric, $\mu_x$ is often called the “median effective dose” (MED), or the “lethal concentration that kills 50%” (LC$_{50}$).

Formulation as a Generalized Linear Model
To formulate this problem as a standard glm, take the response variables to be expressed as observed proportions rather than the counts of expression (16). As we derived in class, the mass functions of these $Y_i$ may be written in exponential dispersion family form as,

$$f(y_i|\theta_i) = \exp[a(\phi)y_i\theta_i - b(\theta_i)] + c(y_i),$$

where

$$\theta_i = \log\left(\frac{p_i}{1-p_i}\right); \quad b(\theta_i) = \log\{1 + \exp(\theta_i)\}; \quad \text{and} \quad \phi \equiv 1; \quad a(\phi) = n_i$$

Now, from (17) $G(\delta_i) = p_i$ so then,

$$G^{-1}(p_i) = \delta_i = \frac{d_i - \mu_x}{\sigma_x} = \frac{-\mu_x}{\sigma_x} + \frac{1}{\sigma_x}d_i,$$

which completes a standard generalized linear model with binomial random component, link function $G^{-1}$, and regression parameters $\beta_0 = -\mu_x/\sigma_x$ and $\beta_1 = 1/\sigma_x$.

Notice from this development that there is a one-to-one relation between distinct tolerance distributions $G(\mu_x, \sigma^2_x)$ and link functions. In particular some of the typical links and tolerance distributions are:
• Normal tolerance distribution, probit link
• Logistic tolerance distribution, logit link
• Extreme value tolerance distribution, complementary log-log link

Now, we would like to fit a model with random component (18), the linear predictor (19) and using the family of link functions,

\[ g(\mu_i | \lambda) = \log \left[ \frac{1}{(1 - \mu_i)^\lambda} - 1 \right] - \log(\lambda). \] (20)

Notice here that we are using \( \mu_i \) for the expected value of \( Y_i \). We have also used \( \mu_x \) for the expected value of the tolerance distribution. It will be important to maintain this distinction in what is to come. Continuing to use \( d_i : i = 1, \ldots, k \) as the experimental doses (i.e., the covariates in the glm), our systematic model component is,

\[ g(\mu_i | \lambda) = \eta_i = \beta_0 + \beta_1 d_i; \quad i = 1, \ldots, k \]

The maximum likelihood algorithm presented earlier can be used to find estimates \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). To do this, however, requires computing a number of quantities in a different manner than would be the case for a model with fixed link. I will not present derivations here, but will list the quantities that would need to be calculated in order to implement the algorithm given in expression (13).

1. Linear Predictor

\[ \eta_i = \beta_0 + \beta_1 d_i \]

2. Means (of \( Y_i \))

\[ \mu_i = 1 - \frac{1}{\left\{ 1 + \lambda \exp(\eta_i) \right\}^{1/\lambda}} \]
3. Derivative of $\eta_i$ wrt $\mu_i$

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{\lambda}{(1 - \mu_i)(1 - (1 - \mu_i)^\lambda)}$$

4. Weights

$$w_i = n_i \left[ \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(\mu_i) \right]^{-1}$$

where $V(\mu_i) = \mu_i(1 - \mu_i)$ as for any model with random component (16)

5. Derivative of $\eta_i$ wrt $\lambda$

$$\frac{\partial \eta_i}{\partial \lambda} = \frac{-\log(1 - \mu_i)}{1 - (1 - \mu_i)^\lambda} - \frac{1}{\lambda}$$

Estimation of Tolerance Distributions

With maximum likelihood estimates of $\beta_0$ and $\beta_1$ in hand we have, from (19) and the invariance property of maximum likelihood, that maximum likelihood estimates of the tolerance function parameters are,

$$\hat{\mu} = \frac{-\hat{\beta}_0}{\hat{\beta}_1}; \quad \hat{\sigma} = \frac{1}{\hat{\beta}_1}$$

(21)

Now, under the theory developed, link function is the inverse distribution function for standardized tolerances (denoted as $\tilde{x}_j$ previously). The distribution function and density function of the tolerances $x_j$ may then be found by inverting expression (20) as follows. We wish to find $G(\tilde{x}) = p$ for some $0 < p < 1$ given that

$$\tilde{x} = G^{-1}(p|\lambda) = \log \left[ \frac{1}{(1 - p)^\lambda} - 1 \right] - \log(\lambda)$$

Then,

$$\frac{1 - (1 - p)^\lambda}{(1 - p)^\lambda \lambda} = \exp(\tilde{x})$$

$$\Rightarrow \quad \exp(\tilde{x}) \lambda(1 - p)^\lambda = 1 - (1 - p)^\lambda$$

$$\Rightarrow \quad (1 - p)^\lambda \{ \lambda \exp(\tilde{x}) + 1 \} = 1$$

$$\Rightarrow \quad p = 1 - \frac{1}{\lambda \exp(\tilde{x}) + 1}^{1/\lambda}$$
So the distribution function for standardized tolerances is then,

\[ G(\tilde{x}|\lambda) = 1 - \frac{1}{\{\lambda \exp(\tilde{x}) + 1\}^{1/\lambda}}. \]  

(22)

Since tolerance \( x_j \) corresponds to standardized tolerance \( \tilde{x}_j \) as \( \tilde{x}_j = (x_j - \mu_x)/\sigma_x \), to obtain the actual distribution function of tolerances, we simply make the appropriate location and scale transformations as

\[ x_j = \sigma_x \tilde{x} + \mu_x \]

and then,

\[ G(x|\lambda, \mu_x, \sigma_x) = \left[ 1 - \frac{1}{\{\lambda \exp\{((x - \mu_x)/\sigma_x) + 1\}^{1/\lambda}} \right] \]  

(23)

which has density

\[ g_t(x|\lambda, \mu_x, \sigma_x) = \frac{\exp\{((x - \mu_x)/\sigma_x)\}}{\sigma_x \{\lambda \exp\{((x - \mu_x)/\sigma_x) + 1\}^{1+1/\lambda}\}} \]  

(24)

An estimated tolerance density then results from substitution of the estimates \( \hat{\mu}_x \) and \( \hat{\sigma}_x \) from (21) into (24).

Example 1 – Bliss Beetle Data

One version of the famous Bliss beetle data is presented in Table 1 (these data have been presented in a number of forms since there were two replicates that are sometimes combined and sometimes not, and concentration is reported on various scales). The data arose from a short-term toxicity test conducted with flour beetles and gaseous carbon disulphide exposure for 5 hours (see Bliss, C.I. (1935), “The calculation of the dosage-mortality curve”, Annals of Applied Biology 22, 134-167).

Estimates of the parameters in the systematic model component were located using a model with a fixed logit link, and a model using the family of links in expression (20). The results are given Table 2. Estimates of \( \mu_x \) and \( \sigma_x \) were obtained from expression (21).
<table>
<thead>
<tr>
<th>Concentration (log)</th>
<th>No. Mortalities</th>
<th>No. Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.893</td>
<td>6</td>
<td>59</td>
</tr>
<tr>
<td>3.970</td>
<td>13</td>
<td>60</td>
</tr>
<tr>
<td>4.042</td>
<td>18</td>
<td>62</td>
</tr>
<tr>
<td>4.108</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>4.171</td>
<td>52</td>
<td>63</td>
</tr>
<tr>
<td>4.230</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>4.285</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>4.338</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 1: Bliss Beetle Data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Logit Link</th>
<th>Link Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-60.640</td>
<td>-39.352</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>14.865</td>
<td>9.518</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>NA</td>
<td>-0.006</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>4.079</td>
<td>4.134</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>0.067</td>
<td>0.105</td>
</tr>
</tbody>
</table>

Table 2: Point Estimates for Bliss Beetle Data
Because the model with a fixed logit link function is nested within the model having an estimated link function (by taking $\lambda = 1$) we may conduct a likelihood ratio test to compare a reduced model (logit link) with a full model (estimated link). The maximized log likelihood (sans constant terms) of the logit link model was $-186.1993$, while that for the model with estimated link was $-182.3464$. This results in a likelihood ratio test statistic $T = -2(-186.1003 + 182.3464) = 7.7058$ and an associated p-value of $p = 0.0055$ (from comparison with a $\chi^2$ distribution having 1 degree of freedom). Thus, we would prefer the model with an estimated link in this example. A plot of the observed responses and fitted curves from both the model with a logit link and the model with an estimated link is presented in Figure 1. A plot of the estimated tolerance density functions is presented in Figure 2.

![Figure 1: Estimated Response Curves for Bliss Beetle Data](image-url)
Our conclusion here would be that there is a clear statistical difference in the abilities of the two models to fit the data, with the model having an estimated link function clearly superior to the model with a fixed logit link.

**Sub-lethal Exposure of Trout to Petroleum Hydrocarbons**

We will apply the same methods used with the Bliss beetle data to another example, this involving the effect of sub-lethal exposure of Rainbow Trout to a petroleum hydrocarbon on response to lethal concentrations of the same substance. The scientific underpinnings of this study involved the fact that petroleum hydrocarbons (e.g., gas and fuel oils) are frequently spilled into aquatic environments. Once released
into the environment, petroleum hydrocarbons bind to sediments and are slowly released back into the water column. Thus, in areas with natural hydrocarbon seeps (near natural deposits), around processing facilities (e.g., offshore oil platforms), or in areas that have been previously polluted, petroleum hydrocarbons are found at the level of $\mu$g/L, which are generally not fatal to fish. On the other hand, “major spills” do also occur with resulting concentrations in the range of mg/L (much higher). The question being investigated was whether a low level of “pre-exposure” makes Rainbow Trout more or less susceptible to higher levels of exposure, should they occur. There is some controversy about whether a low level of chronic exposure makes organisms more resistant or less resistant to particular toxicants.

The study was designed as follows (Steadman et al. (1991), “Decreased Survival of Rainbow Trout Exposed to No. 2 Fuel Oil Caused by Sublethal Preexposure”, Environmental Toxicology and Chemistry 10, 355-363). Test organisms (immature rainbow trout) were exposed to low levels of number 2 fuel oil (2FO) for 21 days. The concentrations used in this “pre-exposure period included 0, 25, and 50 mg/L of 2FO. After 21 days, fish were transferred to other tanks and portions of each pre-exposure group were exposed to 2FO at concentrations of 28.7, 57.4, 114.8, 229.6, 459.1 and 918.2 mg/L for a total of 335 hours. In the actual study, mortality was recorded every 4 or 8 hours and time to death was one of the response variables examined. Here, we will simply consider the data at 258 hours within the context of a typical dose-response analysis. The overall objective is to determine whether the pre-exposure groups differ in their response during the main toxicity test (at least at the time point of 258 hours). We will analyze the data in the same way as for the Bliss beetle data, with a generalized linear model having binomial random component and either logit or estimated link functions. Note that all of the results presented will come from models using log concentration as the covariate.

Parameter estimates for the control treatment (e.g., 0 mg/L of pre-exposure) are
Table 3: Mortalities (Y) and number exposed (N) in a toxicity test with 2FO

<table>
<thead>
<tr>
<th>Pre-exposure concentration</th>
<th>0 mg/L</th>
<th>25 mg/L</th>
<th>50 mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure (mg/L)</td>
<td>Y N</td>
<td>Y N</td>
<td>Y N</td>
</tr>
<tr>
<td>28.7</td>
<td>0 10</td>
<td>0 12</td>
<td>1 9</td>
</tr>
<tr>
<td>57.4</td>
<td>0 10</td>
<td>1 10</td>
<td>0 10</td>
</tr>
<tr>
<td>114.8</td>
<td>2 10</td>
<td>1 10</td>
<td>2 10</td>
</tr>
<tr>
<td>229.6</td>
<td>4 10</td>
<td>4 10</td>
<td>5 10</td>
</tr>
<tr>
<td>459.1</td>
<td>8 9</td>
<td>8 10</td>
<td>7 10</td>
</tr>
<tr>
<td>918.2</td>
<td>10 10</td>
<td>9 9</td>
<td>10 10</td>
</tr>
</tbody>
</table>

Table 4: Parameter estimates for pre-exposure group 0 mg/L

<table>
<thead>
<tr>
<th>Model Link</th>
<th>Logit</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-15.820</td>
<td>4.172</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.905</td>
<td>0.765</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Parameter estimates for the pre-exposure treatment group of 25 mg/L are presented in Table 5 for models with fixed logit and estimated link functions.

Parameter estimates for the pre-exposure treatment group of 50 mg/L are presented in Table 6 for models with fixed logit and estimated link functions.

Because a model with a fixed logit link is nested within a model with link family (20) we may conduct likelihood ratio tests between logit link and estimated link.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Std. Error 1</th>
<th>Estimate 2</th>
<th>Std. Error 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-12.896</td>
<td>3.194</td>
<td>-9.422</td>
<td>2.943</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.338</td>
<td>0.580</td>
<td>1.591</td>
<td>0.553</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>NA</td>
<td>NA</td>
<td>-0.199</td>
<td>0.567</td>
</tr>
</tbody>
</table>

Table 5: Parameter estimates for pre-exposure group 0 mg/L

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Std. Error 1</th>
<th>Estimate 2</th>
<th>Std. Error 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-9.897</td>
<td>2.409</td>
<td>-6.580</td>
<td>2.132</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.816</td>
<td>0.442</td>
<td>1.069</td>
<td>0.404</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>NA</td>
<td>NA</td>
<td>-0.482</td>
<td>0.386</td>
</tr>
</tbody>
</table>

Table 6: Parameter estimates for pre-exposure group 0 mg/L
Maximized Likelihoods

<table>
<thead>
<tr>
<th>Pre-exposure</th>
<th>Logit</th>
<th>Estimated</th>
<th>T</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mg/L</td>
<td>−15.7426</td>
<td>−15.5222</td>
<td>1.0412</td>
<td>0.3075</td>
</tr>
<tr>
<td>25 mg/L</td>
<td>−19.3685</td>
<td>−18.7538</td>
<td>1.229</td>
<td>0.2675</td>
</tr>
<tr>
<td>50 mg/L</td>
<td>−23.7589</td>
<td>−22.7459</td>
<td>2.026</td>
<td>0.1546</td>
</tr>
</tbody>
</table>

Table 7: Likelihoods and LRT tests for logit versus estimated link models.

<table>
<thead>
<tr>
<th>Pre-exposure</th>
<th>$\hat{\mu}_x$</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mg/L</td>
<td>5.685</td>
<td>(5.049, 6.319)</td>
</tr>
<tr>
<td>25 mg/L</td>
<td>5.923</td>
<td>(5.304, 6.543)</td>
</tr>
<tr>
<td>50 mg/L</td>
<td>6.154</td>
<td>(5.303, 7.004)</td>
</tr>
</tbody>
</table>

Table 8: Estimates of tolerance distribution means.

models for each pre-exposure group. Maximized log likelihoods and these likelihood ratio tests are presented in Table 7.

Based on these results we would conclude that there is not sufficient evidence in the data to say that the tolerance distributions in any of the pre-exposure groups differ from a logistic. There is, however, an interesting pattern that suggests itself. The estimated values of the link function parameter $\lambda$ appear to be decreasing as one moves from 0 to 25 to 50 mg/L pre-exposure (Tables 4, 5, 6). Concomitantly, the $p$–values for likelihood ratio tests are becoming smaller as well (Table 7). This might peek our curiosity as to whether anything is being “suggested” by the data in terms of a systematic pattern in the tolerance distributions.

Using results from the models with estimated link functions, values for $\mu_x$ and $\sigma_x$ were arrived at through the use of expression (21) and their standard errors were computed using the delta method in the usual manner. Point and 95% interval estimates of $\mu_x$ are given in Table 8.
These intervals certainly overlap to a great extent, and the same is true for intervals computed under the model with a fixed logit link (not shown). Thus, we are led to the belief that the data do not provide sufficient support for claiming any difference at all between the pre-exposure groups. There is insufficient evidence in the data that pre-exposure of trout to sublethal levels of 2FO changes the response to lethal concentrations at all. What happens if we plot the estimated tolerance distributions for the pre-exposure groups? The estimated densities are presented in Figure 3.

This figure does suggest a systematic change in the tolerance distributions as the level of pre-exposure to 2FO increases from 0 to 25 to 50 mg/L, but is that suggestion one of sensitization (i.e., becoming more susceptible) or acclimation (i.e.,
becoming less susceptible) to the toxicant. Our eye is drawn to the left tails of these estimated densities, which are apparently becoming heavier as the level of pre-exposure increases. But note also the upper portions of the densities which also contain more and more probability as pre-exposure increases. Recall that examination of cumulative densities can often aid in interpretation. The cumulative densities corresponding to the distributions of Figure 3 are presented in Figure 4.

These estimated cumulative densities strongly suggest an effect of acclimation since the rate at which probability (of mortality) is accumulating in these distributions is slower for pre-exposure of 50 mg/L than it is for pre-exposure of 25 mg/L which is in turn slower than for no pre-exposure to 2FO. For example, the cumulative probability in these distributions at a log concentration of 6.0 is 0.862 for the 0 mg/L group, 0.642 for the 25 mg/L group, and 0.432 for the 50 mg/L group.

The overall suggestion from fitting these models is that the effect of pre-exposure, if there in fact is one, is an acclimation effect. Recall that we are not able to conclude there is sufficient evidence in the data to proclaim any effect at all. But consider whether a conclusion that there is, in fact, no effect of pre-exposure is warranted. The study contained small sample sizes (see Table 3) and we might also suggest that in order to detect this type of difference, if indeed one exists, one needs increased information. The point is that, although we cannot conclude from this study that there is any effect of pre-exposure, we are able to suggest (1) if such an effect is real it is likely to be acclimation rather than sensitization, and (2) in order to detect such differences a larger sample size is needed.
Figure 4: Estimated cumulative tolerance densities for Pre-exposure groups