No. 1962
9/1989

MIXTURE MODELS FOR CONTINUOUS DATA IN DOSE-RESPONSE STUDIES WHEN SOME ANIMALS ARE UNAFFECTED BY TREATMENT

Dennis D. Boos and Cavell Brownie

Department of Statistics

North Carolina State University

Raleigh, NC 27695-8203, U.S.A.

SUMMARY

A mixture model is described for dose-response studies where measurements on a continuous variable suggest that some animals are not affected by treatment. The model combines a logistic regression on dose for the probability an animal will "respond" to treatment with a linear regression on dose for the mean of the responders. Maximum likelihood estimation via the EM algorithm is described and likelihood ratio tests are used to distinguish between the full model and meaningful reduced-parameter versions. Use of the model is illustrated with three real-data examples.

Key Words: EM algorithm; Likelihood Ratio tests; Maximum Likelihood; Nonresponse; Normal Mixtures; Outliers.

1. Introduction

In this article we focus on dose-response studies that yield data on a continuous variable and where there appears to be a dose related chance that an individual will not be affected by treatment. Our interest in this "nonresponse" phenomenon developed because of examples encountered in consulting and also in published articles such as Good (1979). Figures 1 and 2 display two data sets that help motivate and illustrate the features of interest here.

Figure 1 presents data from an assay to compare potency of two lots of vaccine. Antibody levels (reported in counts per minute) were determined for mouse sera 21 days after injection and $\log_{10}$ counts are plotted against $\log_{10}$ dose. To reduce crowding, values are offset to the left for lot 1 and to the right for lot 2 at each dose. Figure 2 is a graphical representation of log transformed values for data in Shirley (1977) on reaction times of mice at 4 dose levels of a toxin. Both plots suggest an increase in mean response with increasing dose, but there is considerable noise in the data. Note that the apparent "outliers" follow a pattern which is related to dose. At low doses there are a few large responses while at high doses there are a few noticeably low responses. These extreme values are well separated from the majority of responses in the group and result in a roughly quadratic trend (increasing then decreasing) in variance and in a change in skewness of the response distribution with increasing dose.

The nonresponse phenomenon described by Good (1979) provides a possible explanation for this pattern. Suppose that for each animal there is a threshold that must be exceeded by the administered dose for the animal to show an effect of treatment (i.e., to "respond" to treatment). If the threshold varies among individuals, then the proportion expected to respond to a given dose will be an increasing function of dose. At low doses "responders" appear as outliers, and at high doses the few nonresponders represent the extreme values.

There are various ways to analyze such data. One approach is to test for a dose-related trend in location using a procedure which is valid under the null hypothesis of no dose effect and which has reasonable power in the presence of extreme values (e.g., the Jonckheere-Terpstra test, or see also Shirley, 1977). Such a test does not lead to a quantitative

description of the effect of dose. A second approach is to view the extreme values as atypical observations or outliers and use a robust regression procedure to provide a description of the dose-response relationship. Although robust regression does tend to fit the bulk of the data and to illuminate outliers, the interpretation of fitted values or estimates is not readily related to population means or percentiles. Regression via weighted least squares using a model in which both variance and mean are functions of dose (e.g., Davidian and Carroll, 1987) is another approach that could be considered. All of these methods, however, ignore the information in the pattern of "outliers" caused by the changing proportion of responders, suggesting the need for an analysis which accounts for and utilizes this property of the data. If the pattern of outliers is evidence of an underlying physiological mechanism, then a model motivated by such considerations has the potential to provide information that is more directly interpretable and possibly of a more fundamental nature than that provided by empirical regression approaches. Our goal here is therefore to develop a biologically reasonable model and analysis for data such as those in Figures 1 and 2.

The model we propose is a mixture model like that of Good (1979) but modified to allow a dose-dependent probability of "response" and a dose-dependent magnitude of effect for responders. For simplicity we assume a completely randomized design with n_i animals assigned to receive dose x_i, $i = 1,\ldots,k$ ($n_1+\ldots+n_k = N$). Then for a given dose x_i, the observations Y_{ij}, $j = 1,\ldots,n_i$ are iid with cdf

$$F(y_{ij}|\theta) = p(x_i)G\left(\frac{y_{ij}-\mu-\Delta(x_i)}{\sigma}\right) + (1 - p(x_i))G\left(\frac{y_{ij}-\mu}{\sigma}\right), \tag{1a}$$

where $p(\cdot)$ is a nondecreasing function of x_i between 0 and 1, depending on unknown parameters α and β, μ is the mean for nonresponders, $\Delta(x_i)$ is a function of dose representing the effect for responders and depending on parameters c and d, σ is a scale parameter assumed common to the distributions for responders and nonresponders, and $G(\cdot)$ is a known cdf.

There are numerous choices for specific forms for p, Δ, and G. In implementing the model with several real data sets we have used

$$p(x_i) = (1 + e^{-(\alpha+\beta x_i)})^{-1} \quad ,$$

$$\Delta(x_i) = c + dx_i \ , \tag{1b}$$

and

$$G(y) = \Phi(y), \text{ the standard normal cdf.}$$

In Section 2 we develop the analysis for the model defined by (1a) and (1b), referred to below as the mixture model (1). Theory and some practical issues related to estimation and testing are also described. In Section 3, the mixture model (1) is applied to three data sets to illustrate the types of inferences that are possible. Slight generalizations of the model are introduced for two of the data sets and graphical assessments of model adequacy are provided. Concluding remarks are given in Section 4.

2. Maximum Likelihood Estimation and Hypothesis Testing

2.1 Computation of Estimates and the EM Algorithm

The method of maximum likelihood (ML) has been successfully used in a variety of mixture problems, and both theoretical properties and practical aspects of the method have been well studied (e.g., Titterington, Smith and Makov, 1985; Redner and Walker, 1984). We have therefore relied entirely on ML and likelihood theory for estimation and tests related to the mixture model (1). For computation of estimates Redner and Walker (1984) recommend the EM algorithm for several reasons which include economy of programming and reliable global convergence. We have also found the EM algorithm useful for computations and here briefly outline application of the algorithm for the model in (1).

For $j = 1,\ldots,n_i$, $i = 1,\ldots,k$, let Z_{ij} be an indicator variable that denotes responder status. Thus

$$Z_{ij} = \begin{cases} 1 \text{ with probability } p(\alpha,\beta,x_i) = (1+ e^{-(\alpha+\beta x_i)})^{-1} \\ 0 \text{ with probability } 1 - p(\alpha,\beta,x_i) \end{cases} .$$

The Z_{ij} are not observed (except $Z_{ij} = 0$ may be known for the controls), and the EM algorithm proceeds by treating the observations $Y = (Y_{11},\ldots,Y_{kn_k})'$ as an incomplete data problem, the complete data being (Y, Z). The complete data log likelihood is

$$L_c(Y,Z,\theta) = N \log\left(\frac{1}{\sqrt{2\pi}}\right) - N \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (1-Z_{ij})(Y_{ij}-\mu)^2$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Z_{ij}[Y_{ij} - (\mu + c + dx_i)]^2 \tag{2}$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[(1-Z_{ij})\log(1-p(\alpha,\beta,x_i)) + Z_{ij} \log(p(\alpha,\beta,x_i))\right].$$

The E-step and M-step of the algorithm correspond to calculating updated estimates $\theta^{\nu+1} = (\alpha^{\nu+1},\beta^{\nu+1},\mu^{\nu+1},c^{\nu+1},d^{\nu+1},\sigma^{\nu+1})'$ given current values $\theta^{\nu} = (\alpha^{\nu},\beta^{\nu},\mu^{\nu},c^{\nu},d^{\nu},\sigma^{\nu})'$ by choosing $\theta^{\nu+1}$ as the value of θ which maximizes $Q(\theta,\theta^{\nu}) = E[L_c(Z,Y,\theta)|Y,\theta^{\nu}]$. Here $Q(\theta,\theta^{\nu})$ is L_c of (2) with $\theta = (\alpha,\beta,\mu,c,d,\sigma)$ and with Z_{ij} replaced by the conditional expectation

$$w_{ij}^{\nu} = \frac{p(\alpha^{\nu},\beta^{\nu},x_i)\phi\left(\frac{Y_{ij}-\mu^{\nu}-c^{\nu}-d^{\nu}x_i}{\sigma^{\nu}}\right)\left(\frac{1}{\sigma^{\nu}}\right)}{p(\alpha^{\nu},\beta^{\nu},x_i)\phi\left(\frac{Y_{ij}-\mu^{\nu}-c^{\nu}-d^{\nu}x_i}{\sigma^{\nu}}\right)\frac{1}{\sigma^{\nu}} + (1-p(\alpha^{\nu},\beta^{\nu},x_i))\phi\left(\frac{Y_{ij}-\mu^{\nu}}{\sigma^{\nu}}\right)\frac{1}{\sigma^{\nu}}},$$

$\phi(\cdot)$ being the standard normal pdf. For observations with Z_{ij} unknown, the w_{ij} may be interpreted as posterior probabilites that the Y_{ij} are responders and can be used for identifying the responders. If for any Y_{ij} the value of Z_{ij} is known (e.g., $Z_{ij} = 0$ is often assumed for the controls), then computations are correct if, in Q, w_{ij}^{ν} is set equal to the known value of Z_{ij}.

Note that maximizing $Q(\theta,\theta^{\nu})$ reduces to two separate maximization problems, one involving (α,β) and the other involving (μ,c,d,σ). Explicit solutions for determining $(\mu^{\nu+1},c^{\nu+1},d^{\nu+1},\sigma^{\nu+1})$ are easily derived, but $(\alpha^{\nu+1},\beta^{\nu+1})$ must be determined

numerically as in standard logistic regression (see the Appendix). It is also easily shown that at each iteration there is a unique maximum and hence a unique solution $\theta^{\nu+1}$. Convergence properties of the EM algorithm are summarized in Section 4 of Redner and Walker (1984).

The EM algorithm was programmed and applied to several data sets including those in the examples in Section 3. To check computations, results were compared with estimates and log likelihood (log L) values obtained using PROC NLIN of SAS (1987), with the derivative free option (METHOD = DUD) and with the loss function (_LOSS_) set equal to −log L. Agreement between the two procedures was good but our overall impression is that the EM algorithm is faster, is less sensitive to starting values, and results in smaller values of −log L even with similar convergence criteria.

2.2 Covariance Estimation and Model Adequacy

A method often used to estimate Σ, the covariance matrix of the ML estimates, is to calculate the sample information matrix I(Y), and let $\hat{\Sigma} = I(Y)^{-1}$. It was straightforward (though somewhat tedious) to adapt the method of Louis (1982) to calculate I(Y) for the mixture model (1) (see the Appendix for details). However, since the accuracy of $I(Y)^{-1}$ as an estimator of Σ depends on the model being correct and on sample sizes, we considered two methods for checking the adequacy of $I(Y)^{-1}$.

The first procedure is a simple nonparametric bootstrap: draw iid bootstrap samples $\{Y^*_{ij}, j = 1,\ldots,n_i\}$ with replacement from the data $\{Y_{ij}, j = 1,\ldots,n_i\}$, $i = 1,\ldots,k$. For each of B sets of bootstrap samples, fit the model (1) and obtain $\hat{\theta}^*_1,\ldots,\hat{\theta}^*_B$. Let $\hat{\Sigma}_B$ be the sample covariance of the B bootstrap estimates. $\hat{\Sigma}_B$ will consistently estimate the true covariance matrix of $\hat{\theta}$ regardless of whether the model (1) is correct.

A second method for nonparametrically estimating the covariance matrix of $\hat{\theta}$ follows from an expansion of the log likelihood function log $L(\theta)$,

$$\hat{\theta} - \theta = I(Y)^{-1}\nabla \log L(\theta) + R_n ,$$

where $\nabla \log L(\theta)$ is the gradient of $\log L(\theta)$ and $R_n \xrightarrow{p} 0$ under appropriate regularity conditions. Since the covariance matrix of $\nabla \log L(\theta)$ may be estimated by

$$D(Y) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[\nabla \log f(Y_{ij}|\hat{\theta})\right]\left[\nabla \log f(Y_{ij}|\hat{\theta})\right]',$$

where $f(y|\theta)$is the density of (1a), a nonparametric estimate of the covariance matrix of $\hat{\theta}$ is then $\hat{\Sigma}_D = I(Y)^{-1}D(Y)I(Y)^{-1}$. Similar to the bootstrap estimate $\hat{\Sigma}_B$, $\hat{\Sigma}_D$ will be consistent regardless of whether the model (1) is correct. In fact under iid sampling at each dose $\hat{\Sigma}_B - \hat{\Sigma}_D \xrightarrow{p} 0$, and we have found close agreement between the two estimates in the data from Figure 2 (see Section 3.2 and Table 2). Because $\hat{\Sigma}_D$ is so much simpler to compute than $\hat{\Sigma}_B$, we recommend $\hat{\Sigma}_D$ although it appears to depend more on asymptotic approximation than $\hat{\Sigma}_B$.

Differences between $\hat{\Sigma}_D$ or $\hat{\Sigma}_B$ and $I(Y)^{-1}$ suggest inadequacies of the model (1) or failure of the asymptotics in small samples. We intend to elaborate elsewhere on the comparison of $I(Y)^{-1}$ and $\hat{\Sigma}_D$ for model verification, but in the present article we have relied mainly on plots of fitted means and standard deviations (see Section 3) for that purpose. When the Z_{ij} are known or the component distributions are clearly separated, standard GOF procedures can be used to check the normality assumption. Practically speaking, however, the normality assumption has the advantage that $(\hat{\mu}, \hat{c}, \hat{d}, \hat{\sigma})$ are then explicit and simple to compute in the M step of the EM algorithm.

Within the context of model (1) we test for the adequacy of submodels using standard likelihood ratio methods, i.e., by comparing $-2 \log \Lambda$ to the chi-squared distribution with appropriate degrees of freedom. The usual chi-squared asymptotics will hold as long as $\alpha, \beta \epsilon (-\infty, \infty)$ and no null hypothesis implies that $c = d = 0$ (see Ghosh and Sen, 1985). In particular in our Section 3 examples we use $-2 \log \Lambda$ mainly for tests concerning the regression parameters, i.e., for H_0: $\beta = 0$, H_0: $d = 0$, and H_0: $\beta = d = 0$.

3. Examples

3.1 Comparison of Potency for Two Vaccine Preparations

The data in Figure 1 were generated in a "parallel line" assay to compare the potency of two lots of vaccine. The usual analysis for such data would involve comparing intercepts for the two lots, assuming linear regressions with common slope for the two dose-response functions. In our analysis, we compared the dose-response relationships for the two lots by carrying out hypothesis tests to determine whether any of the parameters of the basic model in (1) were different for the two lots. Preliminary testing indicated d = 0 for both lots, so a 10-parameter model, with α, β, μ, c, and σ assumed lot-specific, was fit to the combined data. With this full model as the alternative, a likelihood ratio test was carried out for each of the null hypotheses H_{01}: $\alpha_1 = \alpha_2$, H_{02}: $\beta_1 = \beta_2$, H_{03}: $\mu_1 = \mu_2$, H_{04}: $c_1 = c_2$, and H_{05}: $\sigma_1 = \sigma_2$, where subscripts on parameters indicate lot dependence. A test of the overall null of no difference between lots (H_{06}: $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\mu_1 = \mu_2$, $c_1 = c_2$, and $\sigma_1 = \sigma_2$) was also carried out.

Table 1 presents values of $-\log L$ for the full and reduced models, and also of $-2 \log \Lambda$ for the likelihood ratio test performed to evaluate each reduced model. The overall test of no difference between lots yields $-2 \log \Lambda = 7.07$ corresponding to an approximate p value of .22 based on the χ^2_5 distribution. Looking at the tests aimed at specific pairs of parameters there is again no indication that α, β, μ, c, or σ differ across lots. We conclude there is no indication of a difference between lots in potency either with respect to the average magnitude of effect for responders, or with respect to the distribution of the response threshold.

Results for Model 6, which assumes no difference between lots, are presented graphically in Figures 3a and 3b. Model-based estimates of the mean and standard deviation as a function of dose were calculated by substituting ML estimates for parameter values in the expressions $E(Y_{ij}) = \mu + p(x_i)\Delta(x_i)$ and $Var(Y_{ij}) = \sigma^2 + p(x_i)(1\text{-}p(x_i))\Delta(x_i)^2$. These fitted curves are

plotted against dose for comparison with the usual sample means and standard deviations calculated separately for each lot. Note that the trend in mean response in Figure 3a is close to linear but there is an indication of a quadratic effect on variance in Figure 3b. Fitted means agree well with the sample means from both lots. The sample standard deviations differ more between lots than do the means, but on average the agreement with the fitted standard deviations is reasonable given the small sample sizes.

3.2 Mouse Reaction Times

The data in Figure 2 are now used to illustrate analysis of a single data set with emphasis on the estimation of estimator precision. Likelihood ratio tests indicated that a reduced model with $d = \beta = 0$ in (1b) was not adequate for the data, but that the model with $d = 0$ was satisfactory ($-\log L$ values were 9.191 for $d = \beta = 0$, 6.590 for $d = 0$, and 6.400 for the unrestricted model). Apparently the proportion of responders increases with dose, but the magnitude of the effect among responders is unaffected by dose, at least for the range of doses studied.

Estimates were therefore calculated under the model with $d = 0$ and model-based means and standard deviations are displayed in Figure 4 together with the usual sample means and standard deviations. Again note the quadratic trend in variability which is nicely fitted by the model. Individual values of the w_{ij} (not displayed) are either close to 0 or close to 1 indicating marked separation between responder and nonresponder distributions. Parameter estimates are presented in Table 2 together with standard errors and correlations obtained using each of the three methods outlined in Section 2.2. From Table 2 we see that the model-based standard errors (computed from $I(Y)^{-1}$) for $\hat{\mu}$, $\hat{c}$ and $\hat{\sigma}$ are very close to the nonparametric estimates from $\hat{\Sigma}_D$ and the bootstrap $\hat{\Sigma}_B$, while the bootstrap standard errors for $\hat{\alpha}$ and $\hat{\beta}$ are larger than for the other two methods. Estimated correlations for the two nonparametric approaches agree closely, but in several instances the nonparametric correlations are substantially different from those calculated using $I(Y)^{-1}$. One possible explanation for this

difference between $\hat{\Sigma}_B$ and $I(Y)^{-1}$ is nonnormality of the responder distribution (e.g., the third moment skewness coefficient $\hat{\gamma}_1 = -.95$ for Y_{ij} values that would be classified as responders using the weights w_{ij}). Overall, as the bootstrap $\hat{\Sigma}_B$ should be the most reliable covariance estimator, it is reassuring that results provided by the computationally simpler $\hat{\Sigma}_D$ and $I(Y)^{-1}$ are similar, with the exception of certain covariances based on $I(Y)^{-1}$.

3.3 Addiction to Morphine in Rats

Weeks and Collins (1971) studied the addiction to morphine in rats in an experiment where, by pressing a lever, rats could obtain morphine by self-injection. There were 9 groups of rats corresponding to 8 concentrations of morphine sulphate solution and a saline control. After 6 days of access to morphine, saline was substituted for the morphine sulphate in each dosed group, and physical dependence was determined for each rat using a weight loss criterion. The number of lever presses (self-injection rates) on the 6th day are displayed in Figure 1 of Good (1979) for the controls and for 4 concentrations of morphine sulphate, with each point identified according to whether the rat was a "responder" (i.e., became physically dependent) or not. These data are reproduced here in Figure 5 as a plot of $Y = \log_{10}(R+1)$ against $x = \log_{10}(\text{dose})$ where R = number of lever presses. Note that there is some overlap of values for responders and nonresponders in the dosed groups and that the proportion of responders is not monotonic on dose. Also, unlike the data in Figures 1 and 2, the means for responders show a regression on dose, the slope of the regression being negative because higher injection rates are needed to deliver the same amount of morphine at low concentrations compared to high concentrations. In practice, the mixture model would not be applied to the data of Figure 5 because responder status (determined by physical dependence) is assumed known for each observation. On the other hand, these data provide an opportunity to assess the mixture model by comparing results with an analysis which utilizes the information on responder status.

As this data set was larger than those in Sections 3.1 and 3.2 (N = 102 compared to 48

and 40, respectively), it was possible to fit a more general model than that in (1). Specifically, the assumption of a common variance for responders and nonresponders was relaxed and (1a) replaced with

$$F(y_{ij}|\theta) = p(x_i)\Phi\left(\frac{y_{ij}-\mu-\Delta(x_i)}{\sigma+\delta}\right) + (1 - p(x_i))\Phi\left(\frac{y_{ij}-\mu}{\sigma}\right) \quad . \tag{3}$$

With (3) as the full model, likelihood ratio tests were then carried out to examine various submodels. Results suggested that the submodel with $\beta = 0$ was adequate for these data, but any other reduced model was too restrictive (e.g., values of $-\log L$ for the full model, the model with $\beta = 0$, and the model with $\beta = \delta = 0$ were 92.446, 92.456, and 97.478, respectively).

The mixture model analysis based on (3) with β=0 leads to several conclusions that can be checked against results based on knowing "responder" status. These are: (i) variability appears to be smaller for responders than for nonresponders ($\hat{\delta} = -.253$), (ii) the mean for responders decreases with the concentration of the morphine solution ($\hat{d} = -.418$), and (iii) there is no evidence of an increase in the proportion of responders (or physically dependent rats) with increasing concentration of the morphine solution ($\beta = 0$). The plots of sample means and standard deviations for responders and nonresponders together with analogous model-based quantities in Figures 6a and 6b support conclusions (i) and (ii). In Figure 6a, there is also a suggestion of an effect of dose on nonresponder means, but the influential point at $x = 0$ is based on one observation only. The actual proportions of responders at log concentrations -1.5, -1.0, $-.5$, and 0 were 11/17, 17/30, 11/17, and 11/12, respectively. Logistic regression of these proportions on log concentrations failed to reject $\beta = 0$ ($-2 \log \lambda = 2.36$, $p > .10$) which agrees with (iii) based on the mixture model analysis. Assuming $\beta = 0$, the proportion of responders for each morphine concentration (i.e., excluding saline controls) was estimated to be .62 and .66, based on the mixture model and logistic regression, respectively.

Model adequacy can also be assessed by comparing the sample means and standard deviations, calculated ignoring responder status, with the corresponding model-based estimates (see Figure 7). Note that the model-based means and standard deviations are not continuous functions of dose, the discontinuity occurring at dose = 0 (arbitrarily plotted in Figure 7 at $x = -2$ log units). The discontinuity is a property of the models (1) or (3) under the assumptions that $\beta = 0$ and that all controls are nonresponders, because then $p(x) = 0$ if $x = 0$ and $p(x) = p > 0$ for $x > 0$. This reduced model is not biologically reasonable but can be regarded as an approximation given a limited range of morphine concentrations. In spite of this limitation, the fitted mean function does seem to describe the pattern in the observed group means. The shape of the fitted standard deviation function seems to agree less well with the concave trend displayed by the group standard deviations. Finally, it is worth noting the following advantage of the mixture model (even with $\beta = 0$) compared to an empirical regression approach. When dose is expressed on a log scale, there is no arbitrariness involved in including the control or dose = 0 group with the mixture model. In contrast, if the controls are to be included, it is not at all obvious how to construct an empirical regression function relating mean response and log dose, for the data of Figure 7.

4. Concluding Remarks

Indiscriminate use of mixture models such as (1) is not recommended (e.g., Farewell, 1986). We suggest the model (1) be employed only when there is good empirical or biological evidence of "nonresponse" or of a similar phenomenon. The separation between groups of data points in Figures 1 and 2 suggests to us the presence of such a mechanism. The model (1) might also be criticized as having unnecessarily many parameters. However, any regression model that adequately accounts for the type of heteroscedascity seen in the example data sets will involve estimation of a comparable number of parameters. For example, a model with linear regression of the mean on dose and quadratic regression of variance on dose involves 5 parameters and does not represent the changes in skewness. Finally, one of our objectives will

be met if this article encourages closer inspection of patterns of variability and outlier occurrence in dose response data, resulting in more critical thought about the underlying biological mechanisms.

Table 1

Values of $-2 \log \Lambda$ for testing the equality across lots of parameters of the mixture model, for the data in Figure 1.

Model	Number of Parameters	$-\log L$	$-2 \log \Lambda$ for reduced vs. full model	p-value
Full Model				
All parameters lot specific (α_i, β_i, μ_i, c_i, σ_i), i=1,2	10	−17.707		
Reduced Models				
Submodels defined by the single constraint:				
1. $\alpha_1 = \alpha_2$	9	−17.366	0.68	.41
2. $\beta_1 = \beta_2$	9	−17.325	0.76	.31
3. $\mu_1 = \mu_2$	9	−17.087	1.24	.27
4. $c_1 = c_2$	9	−17.051	1.31	.25
5. $\sigma_1 = \sigma_2$	9	−16.612	2.19	.14
Completely Reduced Model				
6. All parameters constrained to be equal across lots ($\alpha_1=\alpha_2$, $\beta_1=\beta_2$, $\mu_1=\mu_2$, $c_1=c_2$, $\sigma_1=\sigma_2$)	5	−14.173	7.07	.22

Table 2

Parameter estimates, estimated precision, and correlations under the the mixture model (1) with d=0 for the data in Figure 2.

		Standard Errors		
Parameter	Estimate	$I(Y)^{-1}$	$\hat{\Sigma}_B$	$\hat{\Sigma}_D$
α	−1.90	1.08	1.29	1.08
β	1.12	.53	.65	.53
μ	1.01	.04	.04	.04
c	1.09	.06	.05	.06
σ	0.18	.02	.02	.02

	Correlations		
Estimator Pair	$I(Y)^{-1}$	$\hat{\Sigma}_B$	$\hat{\Sigma}_D$
$\hat{\alpha},\hat{\beta}$	−.93	−.94	−.93
$\hat{\alpha},\hat{\mu}$	−.01	−.16	−.17
$\hat{\alpha},\hat{c}$	−.01	−.02	.09
$\hat{\alpha},\hat{\sigma}$	.01	−.11	−.15
$\hat{\beta},\hat{\mu}$	.01	.16	.12
$\hat{\beta},\hat{c}$	.00	.00	−.06
$\hat{\beta},\hat{\sigma}$	−.01	.06	.10
$\hat{\mu},\hat{c}$	−.64	−.71	−.75
$\hat{\mu},\hat{\sigma}$	−.01	.20	.20
$\hat{c},\hat{\sigma}$	−.01	−.36	−.38

Note. See section 2.2 for the three methods used to calculate standard errors and correlations.

APPENDIX

For the M step of the EM Algorithm given in Section (2.1) we need to maximize $Q(\theta,\theta^{\nu})$, which is just $L_c(Y,Z,\theta)$ with Z_{ij}, replaced by w_{ij}^{ν}. For concise notation we will use p_i for $p(\alpha,\beta,x_i)$, Δ_i for $\Delta(x_i) = c + dx_i$, and omit the ranges for $\Sigma_{i=1}^{k}\,\Sigma_{k=1}^{n_i}$. Recall that if Z_{ij} is known to be 0, then set $p_i = w_{ij} = 0$, and if Z_{ij} is known to be 1, then set $p_i = w_{ij} = 1$. The gradient $\nabla Q(\theta,\theta^{\nu})$ is

$$\frac{\partial Q(\theta,\theta^{\nu})}{\partial \alpha} = \Sigma\Sigma(w_{ij}^{\nu} - p_i) \qquad \frac{\partial Q}{\partial \beta}(\theta,\theta^{\nu}) = \Sigma\Sigma(w_{ij}^{\nu} - p_i)x_i$$

$$\frac{\partial Q(\theta,\theta^{\nu})}{\partial \mu} = \Sigma\Sigma\Big[(1 - w_{ij}^{\nu})(Y_{ij}-\mu) + w_{ij}^{\nu}(Y_{ij}-\Delta_i)\Big]\Big/\sigma^2$$

$$\frac{\partial Q(\theta,\theta^{\nu})}{\partial c} = \Sigma\Sigma\Big[w_{ij}^{\nu}(Y_{ij} - \Delta_i)\Big]\Big/\sigma^2$$

$$\frac{\partial Q(\theta,\theta^{\nu})}{\partial d} = \Sigma\Sigma\Big[w_{ij}^{\nu}(Y_{ij} - \Delta_i)\Big]x_i\Big/\sigma^2$$

$$\frac{\partial Q(\theta,\theta^{\nu})}{\partial \sigma} = \Sigma\Sigma\Big[-\sigma^2 + (1-w_{ij}^{\nu})(Y_{ij}-\mu)^2 + w_{ij}^{\nu}(Y_{ij}-\Delta_i)^2\Big]\Big/\sigma^3 \ .$$

The Hessian $H(\theta)$ of $Q(\theta,\theta^{\nu})$ is block diagonal with the (α,β) part similar to that for standard logistic regressions: the (θ_k,θ_l) elements of $H(\theta)$ are $\Sigma\Sigma h_{ij}(\theta_k,\theta_l;\theta)$, where $h_{ij}(\alpha,\alpha;\theta) = -(1-p_i)p_i$, $h_{ij}(\alpha,\beta;\theta) = -(1-p_i)p_ix_i$, and $h_{ij}(\beta,\beta;\theta) = -(1-p_i)p_ix_i^2$. The estimates $(\hat{\alpha}^{\nu+1},\hat{\beta}^{\nu+1})$ are unique and can be found easily by Newton-Raphson iteration of $(\hat{\alpha}^{p+1}, \hat{\beta}^{p+1})' = (\hat{\alpha}^{p},\hat{\beta}^{p})' - H_{22}(\hat{\alpha}^{p},\hat{\beta}^{p})^{-1}\nabla Q(\theta^{p},\theta^{\nu})$, where $(\hat{\alpha}^{\nu},\hat{\beta}^{\nu})$ are used as starting values and $H_{22}(\alpha,\beta)$ is the 2×2 part of $H(\theta)$ relating to (α,β).

The estimates $(\hat{\mu}^{\nu+1}, \hat{c}^{\nu+1}, \hat{d}^{\nu+1}, \hat{\sigma}^{\nu+1})$ are explicitly found by setting $\nabla Q(\theta,\theta^{\nu}) = 0$; e.g.,

$$\hat{\mu}^{\nu+1} = \left[\Sigma\Sigma Y_{ij} - \Sigma\Sigma w_{ij}^{\nu} Y_{ij}\right] \Big/ \left[N - \Sigma\Sigma w_{ij}^{\nu}\right],$$

$$\hat{\sigma}^{\nu+1} = \frac{1}{N}\,\Sigma\Sigma\left[(1-w_{ij}^{\nu})(Y_{ij}-\hat{\mu}^{\nu+1})^2 + w_{ij}(Y_{ij} - \hat{\Delta}_i^{\nu+1})^2\right].$$

Following Louis (1982), the sample information matrix I(Y) (see Section 2.2) may be computed as $I(Y) = -H(\hat{\theta}) - K(\hat{\theta})$, where $\hat{\theta}$ is the ML estimate and

$$K(\theta) = E\left[\nabla L_c(Y,Z,\theta) \cdot (\nabla L_c(Y,Z,\theta))' | Y,\hat{\theta}\right].$$

The elements of $H(\hat{\theta})$ relating to (α,β) were given above. The remaining nonzero elements are $H_{\hat{\mu}\hat{\mu}} = -N/\hat{\sigma}^2$, $H_{\hat{\mu}\hat{c}} = H_{\hat{c}\hat{c}} = -\Sigma\Sigma\hat{w}_{ij}/\sigma^2$, $H_{\hat{\mu}\hat{d}} = H_{\hat{c}\hat{d}} = -\Sigma\Sigma\hat{w}_{ij}x_i/\hat{\sigma}^2$, $H_{\hat{d}\hat{d}} = -\Sigma\Sigma\hat{w}_{ij}x_i^2/\hat{\sigma}^2$, and $H_{\hat{\sigma}\hat{\sigma}} = -2N/\hat{\sigma}^2$, where $\hat{w}_{ij}$ is w_{ij}^{ν} evaluated at $\theta^{\nu} = \hat{\theta}$.

The (θ_k,θ_l) elements of $K(\hat{\theta})$ are $\Sigma\Sigma(1-\hat{w}_{ij})\hat{w}_{ij}a_{ij}(\hat{\theta}_k,\hat{\theta}_l)$, where $a_{ij}(\hat{\alpha},\hat{\alpha}) = 1$, $a_{ij}(\hat{\alpha},\hat{\beta}) = x_i$, $a_{ij}(\hat{\beta},\hat{\beta}) = x_i^2$, $a_{ij}(\hat{\mu},\hat{\alpha}) = -\hat{\Delta}_i/\hat{\sigma}^2$, $a_{ij}(\hat{\mu},\hat{\beta}) = -\hat{\Delta}_i x_i/\hat{\sigma}^2$, $a_{ij}(\hat{\mu},\hat{\mu}) = \hat{\Delta}_i^2/\hat{\sigma}^4$, $a_{ij}(\hat{c},\hat{\alpha}) = (Y_{ij}-\hat{\Delta}_i)/\hat{\sigma}^2$, $a_{ij}(\hat{c},\hat{\beta}) = (Y_{ij}-\hat{\Delta}_i)x_i/\hat{\sigma}^2$, $a_{ij}(\hat{c},\hat{\mu}) = -\hat{\Delta}_i(Y_{ij}-\hat{\Delta}_i)/\hat{\sigma}^4$, $a_{ij}(\hat{c},\hat{c}) = (Y_{ij}-\hat{\Delta}_i)^2/\hat{\sigma}^4$, $a_{ij}(\hat{d},\hat{\alpha}) = (Y_{ij}-\hat{\Delta}_i)x_i/\hat{\sigma}^2$, $a_{ij}(\hat{d},\hat{\beta}) = (Y_{ij}-\hat{\Delta})x_i^2/\hat{\sigma}^2$, $a_{ij}(\hat{d},\hat{\mu}) = -\hat{\Delta}_i(Y_{ij}-\hat{\Delta}_i)x_i/\hat{\sigma}^4$, $a_{ij}(\hat{d},\hat{c}) = (Y_{ij}-\hat{\Delta}_i)^2x_i/\hat{\sigma}^4$, $a_{ij}(\hat{d},\hat{d}) = (Y_{ij}-\hat{\Delta}_i)^2x_i/\hat{\sigma}^4$, $a_{ij}(\hat{\sigma},\hat{\alpha}) = [(Y_{ij}-\hat{\Delta}_i)^2-(Y_{ij}-\hat{\mu})^2]/\hat{\sigma}^3$, $a_{ij}(\hat{\sigma},\hat{\beta}) = a_{ij}(\hat{\sigma},\hat{\alpha})x_i$, $a_{ij}(\hat{\sigma},\hat{\mu}) = -a_{ij}(\hat{\sigma},\hat{\alpha})\hat{\Delta}_i/\hat{\sigma}^2$, $a_{ij}(\hat{\sigma},\hat{c}) = a_{ij}(\hat{\sigma},\hat{c}) = a_{ij}(\hat{\sigma},\hat{\alpha})(Y_{ij}-\hat{\Delta}_i)/\hat{\sigma}^2$, $a_{ij}(\hat{\sigma},\hat{d}) = a_{ij}(\hat{\sigma},\hat{c})x_i$, and $a_{ij}(\hat{\sigma},\hat{\sigma}) = [(Y_{ij}-\hat{\Delta}_i)^2-(Y_{ij}-\hat{\mu})^2]^2/\hat{\sigma}^6$.

REFERENCES

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* 82, 1079-1091.

Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *The Canadian Journal of Statistics* 14, 257-262.

Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, L. M. LeCam and R. A. Olshen (eds.), Vol. 2, 789-806, Pacific Grove, CA: Wadsworth.

Good, P. I. (1979). Detection of a treatment effect when not all experimental subjects will respond to treatment. *Biometrics* 35, 483-489.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* 44, 226-233.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* 26, 195-239.

SAS Institute Inc. (1987). *SAS/STAT Guide for Personal Computers, Version* 6 *Edition*, Cary, NC: SAS Institute Inc.

Shirley, E. A. C. (1977). A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics* 33, 386-389.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley.

Weeks, J. R. and Collins, R. J. (1971). Primary addiction to morphine in rats. *Federation Proceedings* 30, 277 ABS.

Figure Captions

Figure 1. Mouse serum antibody levels, as $\log_{10}$ (count per minute) or log CPM, 21 days after injection with vaccine, for two lots of vaccine at 3 doses. At each dose, values are offset to the left for lot 1 (◆) and to the right for lot 2 (Δ).

Figure 2. Log reaction times of mice, by dose, from Shirley (1977). Doses reported as low, medium, and high in Shirley (1977) are treated here as equally spaced.

Figure 3a. Sample means by lot (◆ and Δ) and fitted means (−) for serum antibody levels of Figure 1. Fitted means are based on the mixture model (1) with d = 0 and assuming no differences between lots.

Figure 3b. Sample standard deviations by lot (◆ and Δ) and fitted standard deviations (−) for serum antibody levels of Figure 1.

Figure 4. Sample (Δ) and fitted (0) means and standard deviations for mouse reaction times of Figure 2. Fitted values are based on the mixture model (1) with d = 0.

Figure 5. Self-injection rates, as $\log_{10}$ (# of lever presses + 1) or log (count + 1), for rats after 6 days of exposure to morphine. Physically dependent rats or responders (◆), identified by a weight loss criterion, are distinguished from nonresponders (Δ). Data from Good (1979).

Figure 6a. Sample and fitted means for responders (◆ and −) and nonresponders (Δ and −) for self-injection rates of Figure 5. Fitted values are based on the mixture model (3) with $\beta = 0$.

Figure 6b. Sample and fitted standard deviations for responders (⊕ and $-$) and nonresponders (Δ and $-$) for self-injection rates of Figure 5.

Figure 7. Sample means (0) and standard deviations ($*$) together with fitted means (◆ for log dose $=-\infty$ and —— for log dose > -1.5) and fitted standard deviations (▲ for log dose $=-\infty$ and —— for log dose > -1.5), for self injection rates of Figure 5. Fitted values are based on the mixture model (3) with $\beta = 0$.

Figure 1

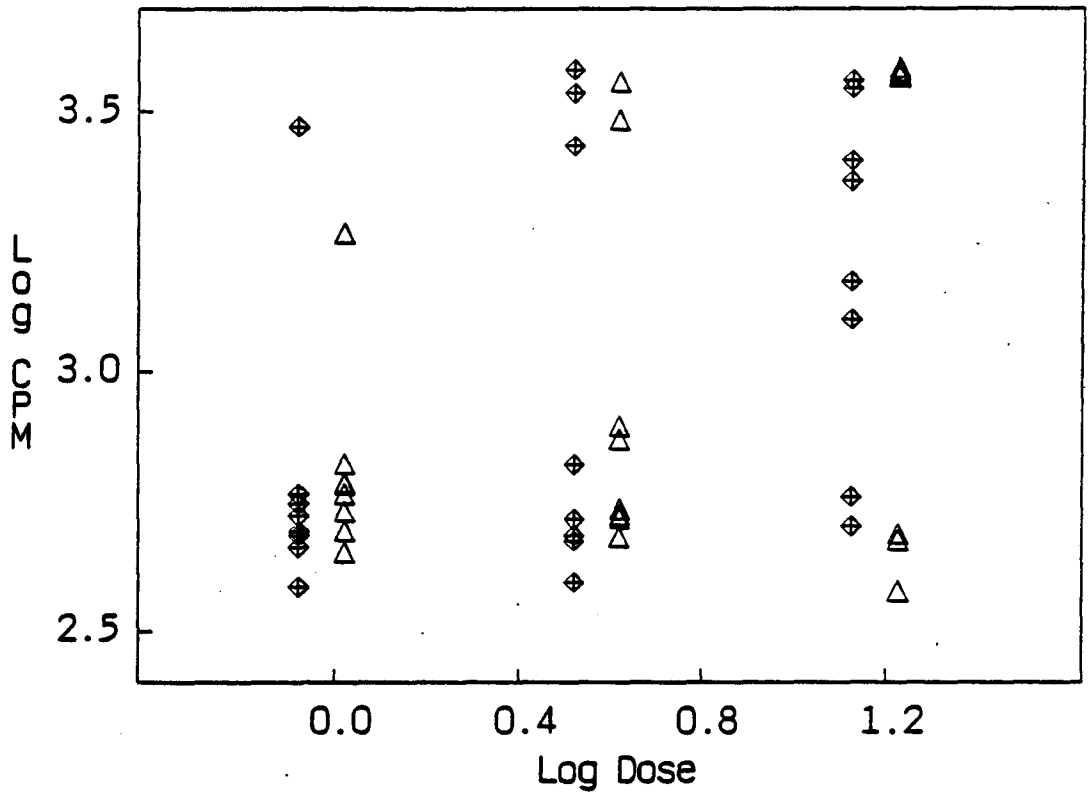


Figure 2

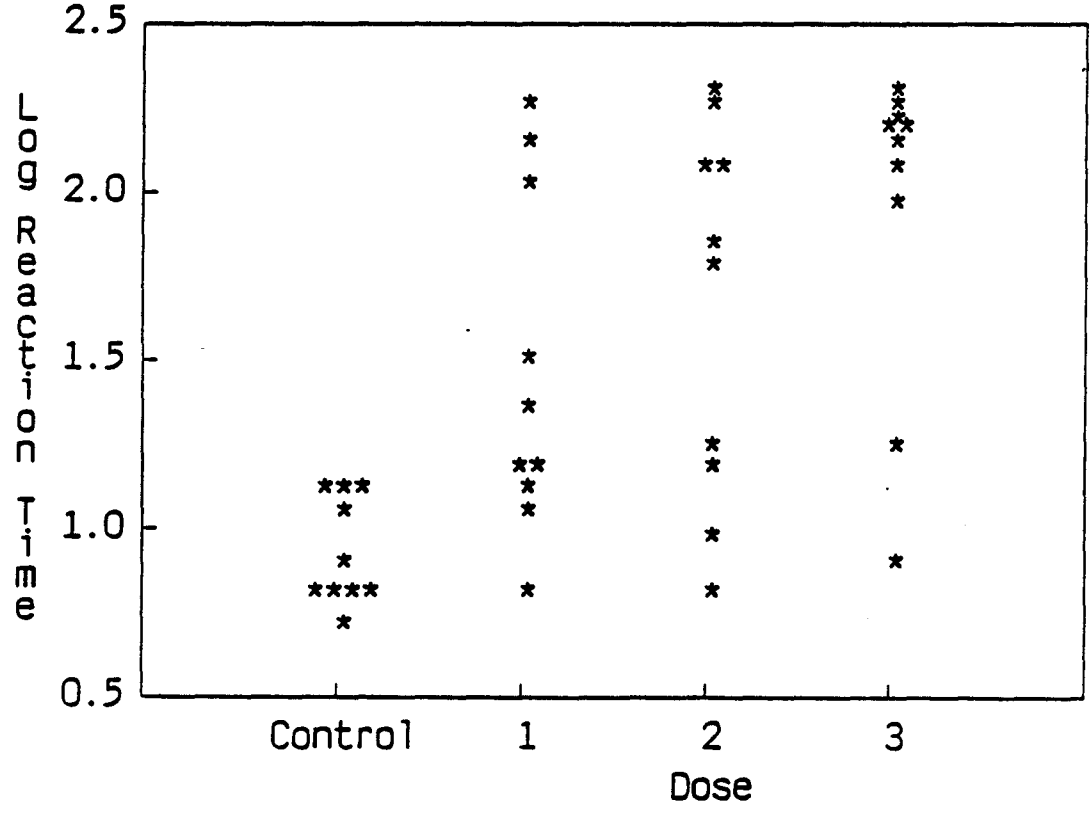


Figure 3A

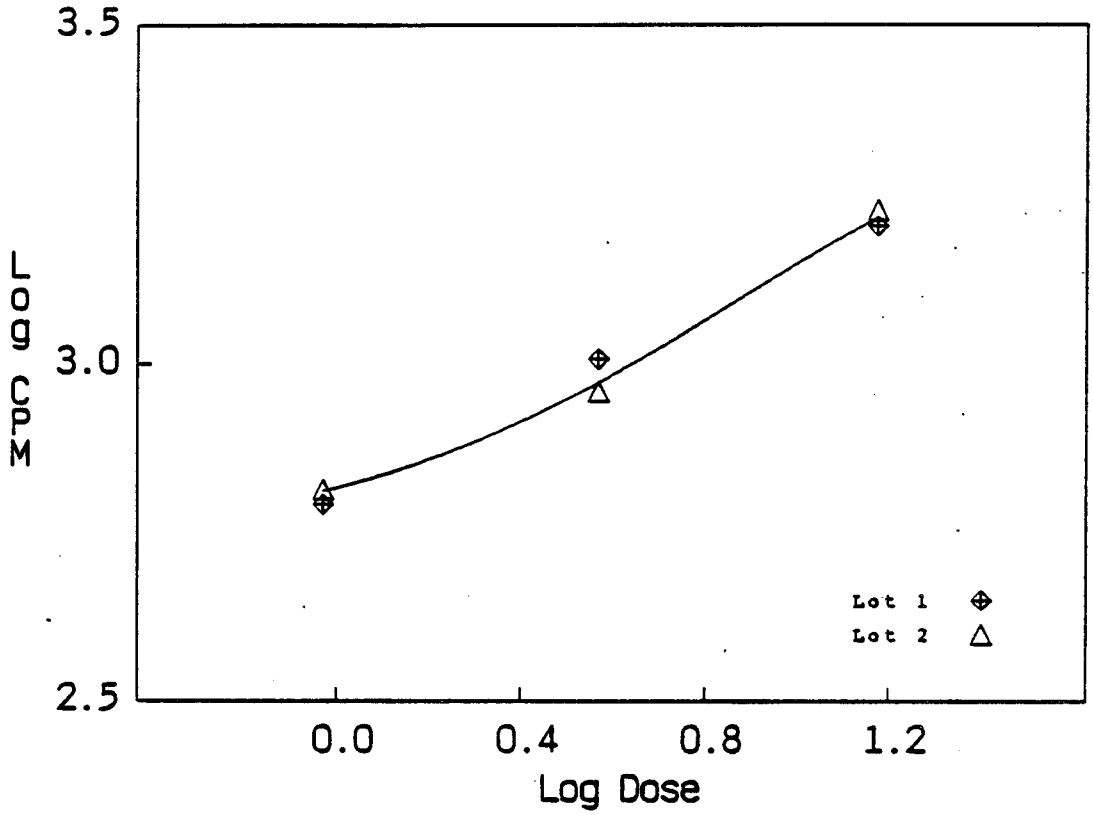


Figure 3B

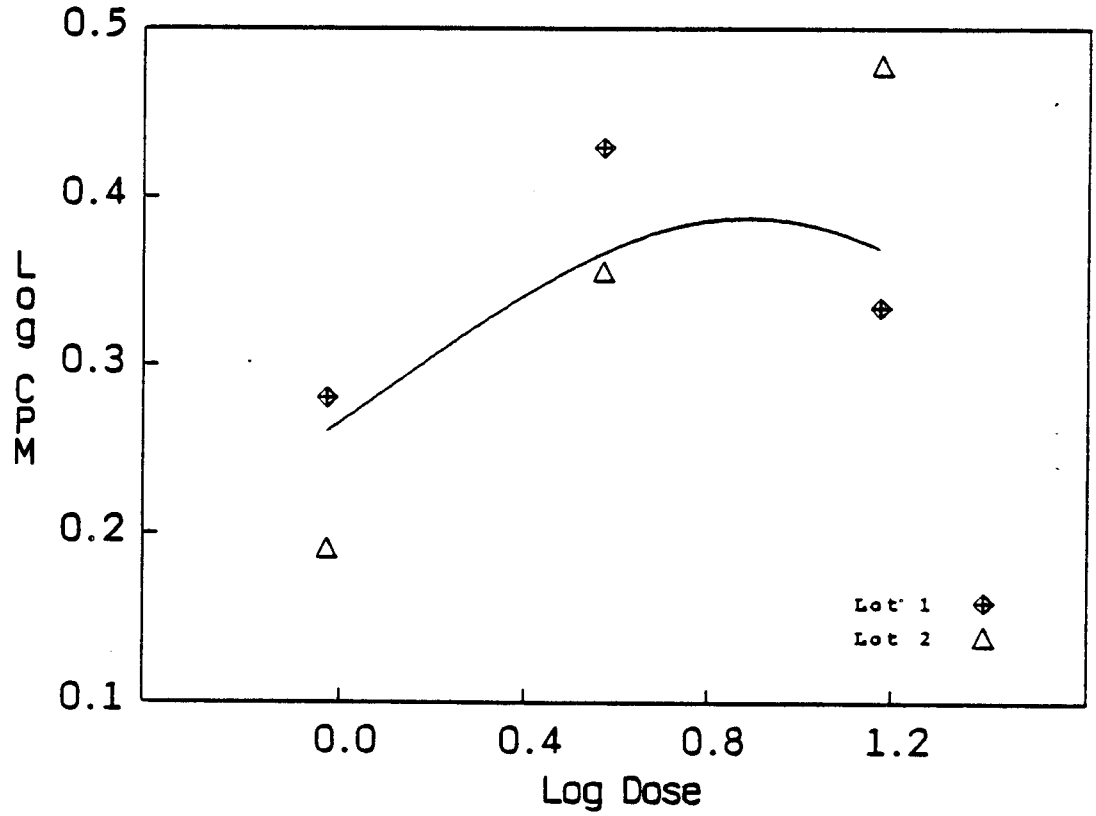


Figure 4

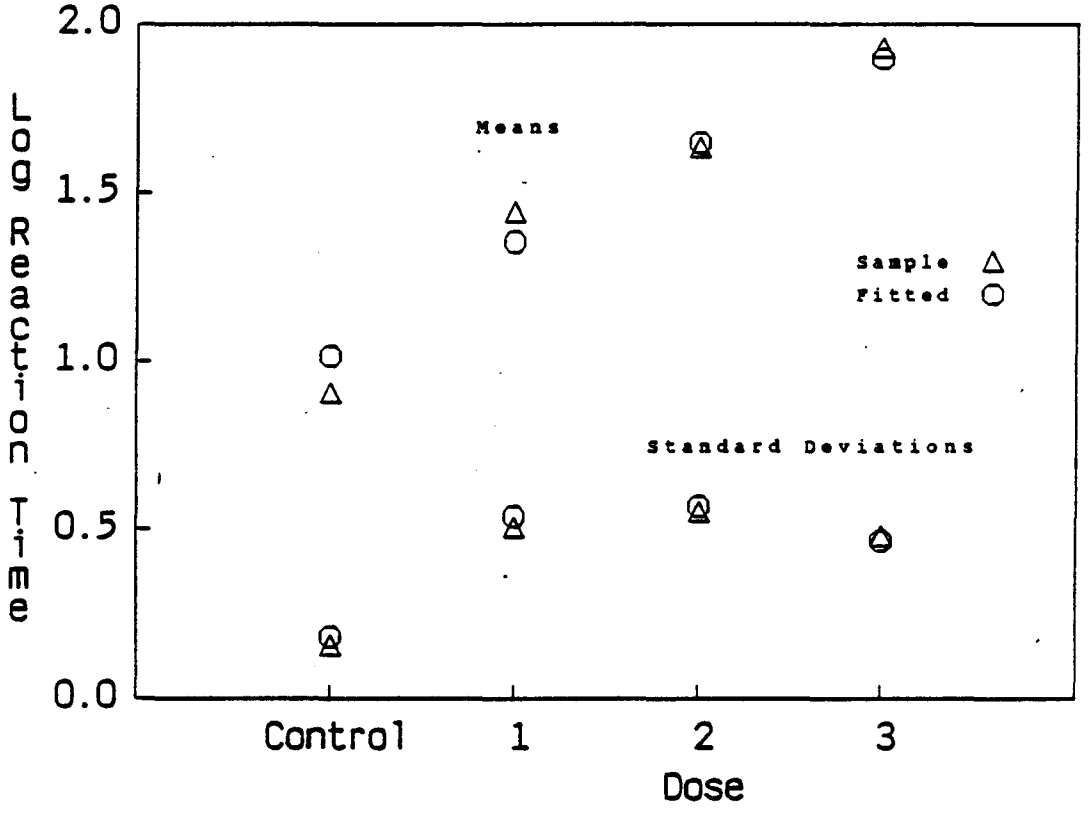

2.0
1.5
1.0
0.5
0.0
Log Reaction Time
Means
Sample
Fitted
Standard Deviations
Control
1
2
3
Dose

Figure 5

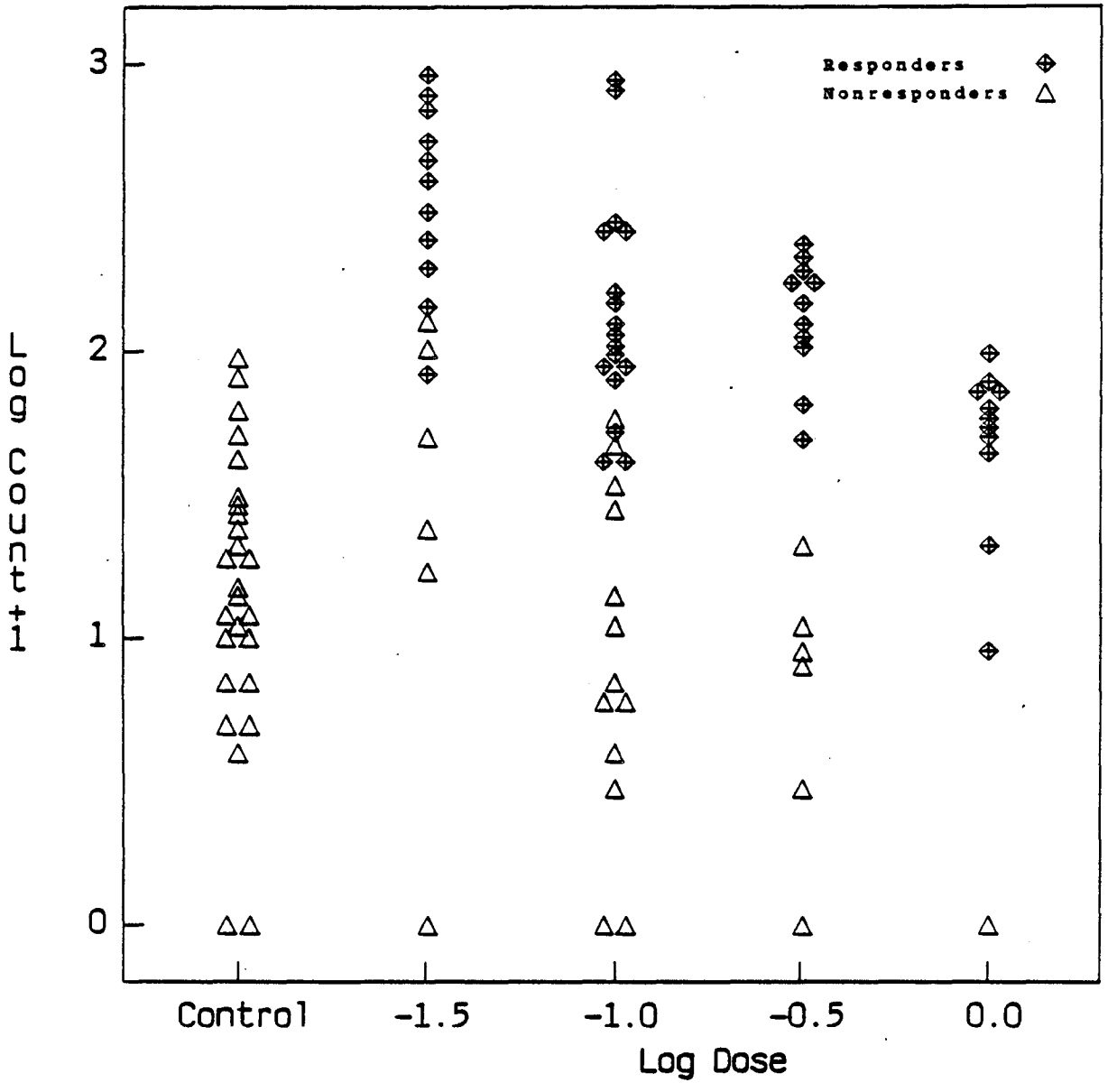

Responders
Nonresponders
Log Count+1
3
2
1
0
Control
-1.5
-1.0
-0.5
0.0
Log Dose

Figure 6A

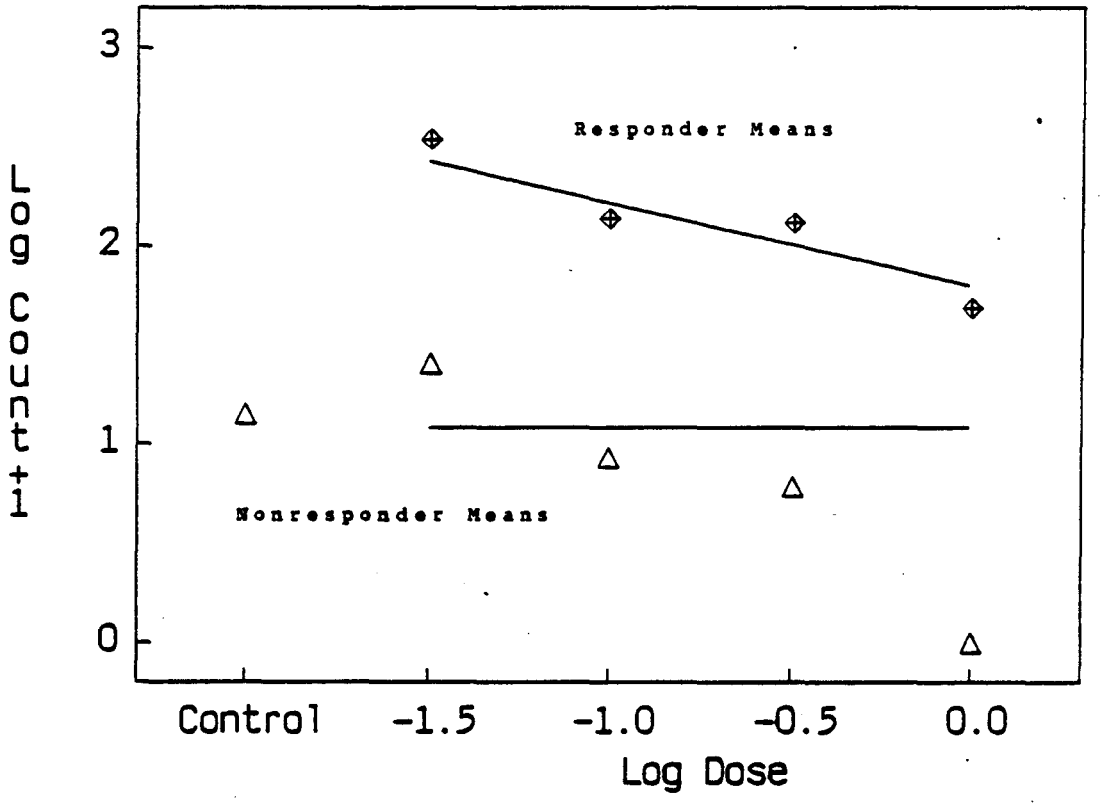


Figure 6B

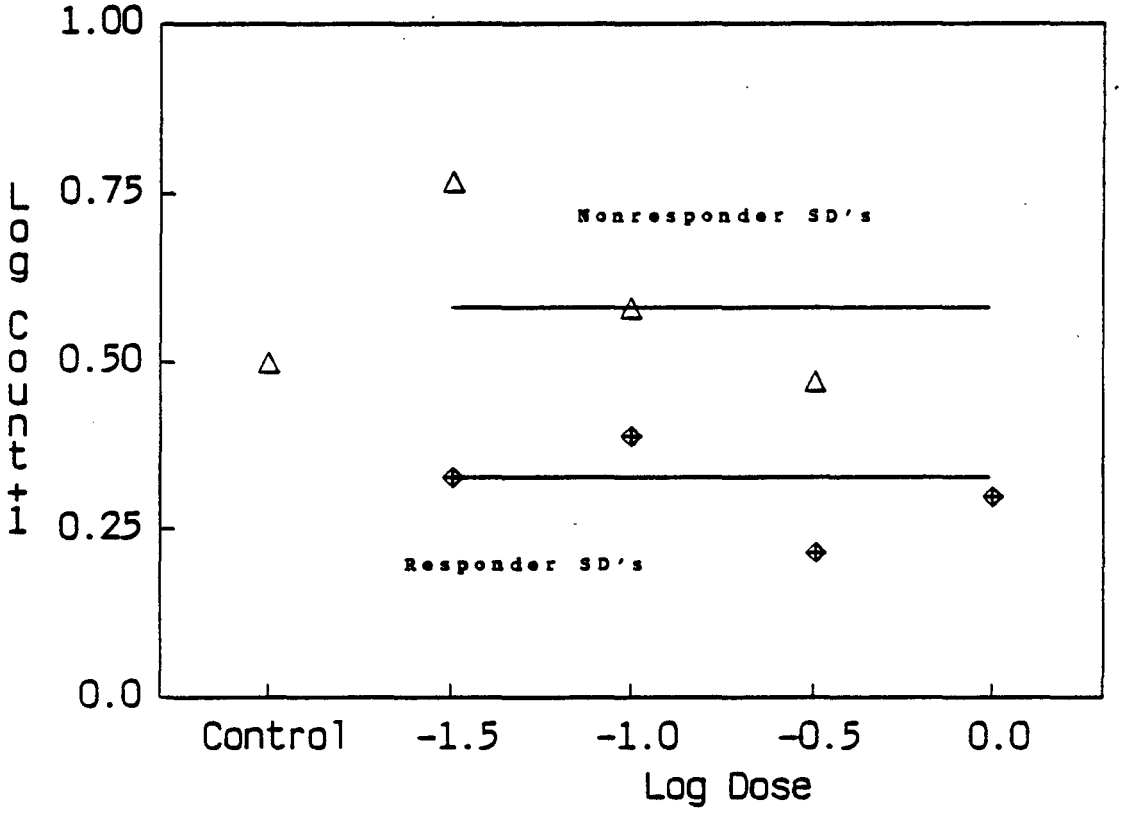
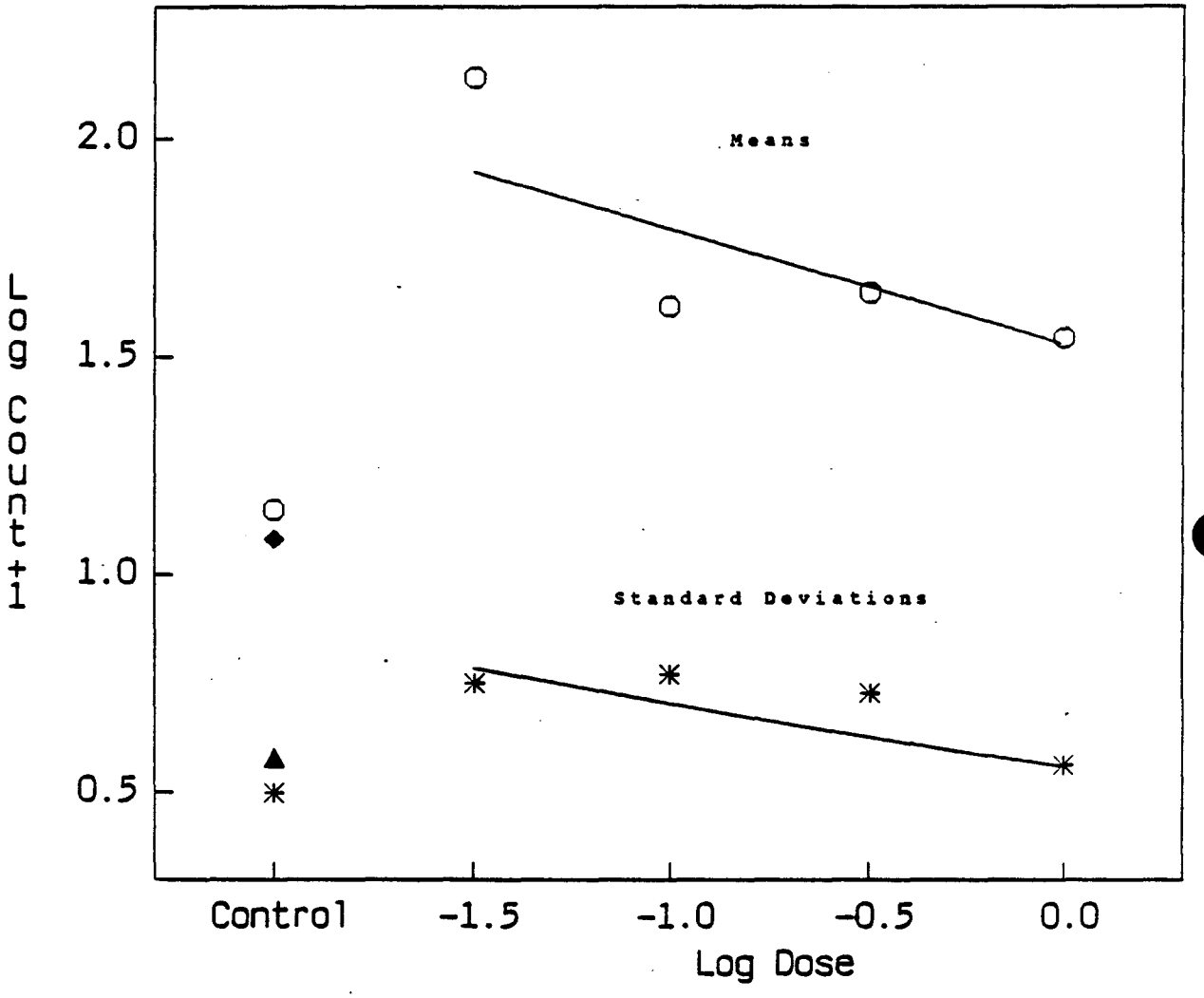


Figure 7


Means
Standard Deviations
Log Count+1
2.0
1.5
1.0
0.5
Control
-1.5
-1.0
-0.5
0.0
Log Dose