A NONLINEAR REGRESSION APPROACH TO ESTIMATING SIGNAL DETECTION MODELS FOR RATING DATA

CHING-FAN SHEU ANDREW HEATHCOTE

DEPAUL UNIVERSITY UNIVERSITY OF NEWCASTLE

Key words: signal detection models, nonlinear regression.

Address correspondence to Ching-Fan Sheu, Department of Psychology, DePaul University, 2219 N. Kenmore Ave, Chicago, Illinois 60614-3522 (e-mail:csheu@dePaul.edu).
This paper considers a regression approach to estimating signal detection parameters for rating data. The methodology is based on the statistical modeling of ordinal data and requires only standard statistical software such as SAS for computation. The approach is more efficient than the current practice of extracting the parameter estimates using specialized software and analyzing the estimates using a standard statistical package. It greatly facilitates exploration of the effects of covariates on model parameters. The method is illustrated using a published data set from a single factor multiple-alternative perceptual task, and data from a more complex factorial design examining recognition memory rating data.
Introduction

This paper presents a method for fitting covariate-adjusted signal detection models to rating data from discrimination experiments. In the regression framework, treatment conditions can be coded as indicator covariate variables so that the effects of experimental factors on parameter estimates of the signal detection model can be assessed directly from the regression coefficients of a generalized (ordinal) regression equation.

The ability to directly analyze effects of covariates (factors in experimental designs) on parameters provides a flexible approach to data analysis using theory of signal detectability (MacMillan & Creelman, 1991). For example, it allows simultaneous fitting of all conditions in a complex experiment rather than comparison of results from multiple hit versus false alarm ROC analyses. It also allows a nested (full versus reduced) model comparison to be conducted without having to rewrite the software for the estimating procedure when a different reduced (non-saturated) model is considered.

A widely used approach to fitting the confidence ratings resulting from a discrimination task is to assume a (latent) normal model (Green & Swets, 1974) and to estimate the parameters separately for each pair of ratings from signal and noise conditions using a maximum likelihood method (Dorfman & Alf, 1969). Kijewski, Swensson, and Judy (1989) extended this approach to simultaneously fit rating data from more than two alternatives. Their fitting algorithm requires second-order derivative equations for the maximum-likelihood estimation. Their work illustrates the difficulties an experimenter faces in applying signal detection models for data collected from factorial experimental designs. The experimenter must be able to differentiate with respect to the unknown parameters in the maximum-likelihood estimating equations and implement the calculations in a computer language. Such an approach
is not practical as it demands the construction of a specialized procedure for each experimental design. The approach is also inefficient. The researcher typically wishes to assess the effect of a treatment condition on the parameter estimate, such as the difference between signal and noise means. The standard approach requires that this is accomplished in two stages. First the parameter estimates is extracted from noise and signal conditions of an experiment using some specialized computer programs, and then a standard statistical procedure such as analysis of variance is performed on the estimates as if they were raw data. A more efficient approach would allow the effects of factors in an experimental design on parameter estimates to be assessed directly.

Tosteson and Begg (1988) proposed a regression approach to estimate receiver-operating-characteristic curves in radiology. Their ordinal regression methodology is built on the framework of the generalized linear models (McCullagh & Nelder, 1989) and allows the estimation of location and scale parameters to be adjusted for relevant covariates. Tosteson and Begg used a specialized computer program for their analysis. DeCarlo (1998) discussed the connection between signal detection theory and generalized linear models and presented sample SAS and SPSS programs for well known data sets in signal detection literature. Unfortunately, the often used unequal variance normal model is not of the generalized linear type, and there is no built-in procedure in general-purpose statistical software to estimate such model parameters automatically.

Some researchers have estimated the parameters of the unequal variance signal detection model using multiple linear regression on transformed cumulative ratings proportions (e.g. Ratcliff, McKoon & Tindall, 1994). The transformation used depends on the underlying form of the latent distribution. This approach has several weaknesses. First, ratings from all but one condition are used as predictors, and so
are assumed to be measured without error, a clearly wrong assumption. Further, error is minimized by regression on the remaining condition, giving it undue influence. Finally, each data point is equally weighted but clearly some have much greater estimation variance than others. For reasonably accurate responding, for example, there are usually few high confidence errors and so data points for these ratings are very variable. The problem is exacerbated by the strongly nonlinear effect of the transformation when cumulative probabilities are near zero or one. As a consequence, this approach is likely not to be robust. However, none of these weaknesses of an ad-hoc linear regression approach applies to maximum likelihood estimation. It is well known that maximum likelihood estimation can be achieved by means of nonlinear least squares (Jennrich & Moore, 1975). In this paper, we follow the method suggested by Cox (1984), who treated the problem of maximizing likelihood as a regression problem and used non-linear regression programs, which are available in standard statistical packages, to estimate signal detection parameters.

We believe that the requirement of using specialized software prohibits the regression methodology from being widely applied by psychologists for the analysis of signal detection experiments. A goal of this paper is to bridge the gap between standard regression modeling of ordinal data and the practice of ROC data analysis in psychology. We acknowledge that the signal detection models, being a class of the latent variable models, could also be formulated within the structural equation modeling framework (Bollen, 1989) with the estimation procedure accomplished by LISREL (Jöreskog & Sörbom, 1993). The regression approach is, however, likely to be more familiar to psychologists and is readily applicable to experimental data. Furthermore, the nonlinear regression methodology allows reasonable non-normal models such as logistic or complementary log-log distributions to be implemented with minimal changes in program code.
In this article, we first briefly review the signal detection model for rating data. The second section illustrates Tosteson and Begg’s (1988) approach by formulating a regression model for Kijewski et al’s (1989) data from a multiple-alternative perceptual task. The third section describes the details of using the SAS non-linear regression program for nonlinear model fitting. The last section presents an application of the regression approach to confidence rating data from a recognition memory experiment.

The Signal Detection Model for Rating-ROC Data

This section illustrate the construction of signal detection models for rating data and introduces notation. Subjects in a discrimination experiment are asked to use a fixed number of categories to indicate how confident they are that a signal has been presented in an experimental trial. For example, four response categories from 1 = ‘noise for sure’ to 4 = ‘signal for sure’ may be used. A normal model for this case assumes a normal distribution for noise, with mean $\mu_1$ and standard deviation $\sigma_1$, and a normal distribution for signal, with mean $\mu_2$ and standard deviation $\sigma_2$. To fit four categorical responses, 3 (latent) cutpoints are assumed. Thus, depending on whether a noise or a signal stimulus has been presented, we have ($\Phi$ denotes the cumulative distribution function of a standard normal random variable):

$$\Pr(R = 1|S_i) = \Phi((c_1 - \mu_i)/\sigma_i);$$
$$\Pr(R = 2|S_i) = \Phi((c_2 - \mu_i)/\sigma_i) - \Phi((c_1 - \mu_i)/\sigma_i);$$
$$\Pr(R = 3|S_i) = \Phi((c_3 - \mu_i)/\sigma_i) - \Phi((c_2 - \mu_i)/\sigma_i);$$
$$\Pr(R = 4|S_i) = 1 - \Phi((c_3 - \mu_i)/\sigma_i);$$

where $R$ is the categorical response for a particular trial in the experiment, and $c_1$, $c_2$, $c_3$ are the cutpoints (decision thresholds). A subject’s categorical responses are used to estimate the location-shift parameter ($\mu_2 - \mu_1$), the scale parameter ($\sigma_1/\sigma_2$),
and the cutpoints. Typically, $\mu_1$ is set to zero and $\sigma_1$ to one for identifiability. This is equivalent to choosing the noise distribution as the reference for comparison. Note that if the noise and signal distributions have the same variance, then the scale parameter reduces to a constant of one. This is the equal-variance normal model and it belongs to the class of generalized linear models. In the case where noise and signal distributions have different variances, the constraint that the scale parameter has to be positive excludes the unequal-variance normal model from the family of the generalized linear models. Hence, the unequal-variance model must be estimated using the maximum-likelihood method presented in Dorfman and Alf (1969) or nonlinear regression procedures such as that proposed by Cox (1984).

The Regression Approach to ROC Data

Tosteson and Begg (1988) proposed a generalization of the signal detection models, allowing the shift in location and/or scaling to depend on each of the distributions (noise or signal) under consideration. This family of models also appeared on page 154 of McCullagh and Nelder (1989). We present a regression model of this type with a single continuous covariate for data from a perceptual experiment reported in Kijewski et al (1989). One observer’s 12-category rating of 196 visual stimuli were recorded under five levels of stimulus contrast: -2, -1, 0, 1, and 2. The number of stimulus presentations (trials) for each stimulus contrast was, 38, 38, 40, 40, and 38, respectively. The model proposed by Kijewski et al’s (1989) has a total of 19 parameters: a pair of location and scale parameters for each of the four distributions corresponding to non-zero contrast levels plus 11 cutpoint parameters. We noticed that the parameter estimates reported in Kijewski et al’s (1989) paper were linearly related to the stimulus contrast levels, so we propose a simple model with one location parameter and one scale parameter that determine the mean and standard
deviation of each condition as a linear function of contrast.

Let \( \pi_{ij} \) be the probability of jth categorical response \( (j = 1, 2, \ldots, 12) \) for distribution \( i \) \( (i = 1, \ldots, 5) \). For example, \( \pi_{36} \) is the (theoretical) probability of responding with a judgment of category 6 when the level of stimulus contrast is zero. The observed frequency of this categorical response was 14. These multinomial probabilities may be determined by a normal model for 12 categories and 5 distributions similar to the normal model described earlier for 4 categories and 2 distributions. We will assume that the distribution for the zero contrast \( (i = 3) \) condition is used as reference with a mean zero and variance of one. The model we consider has a total of 13 parameters and takes the following form:

\[
p_{ij} = \Phi\left(\frac{c_i - \beta x_i}{\exp(\alpha x_i)}\right),
\]

(1)

where \( p_{ij} = \pi_{1j} + \cdots + \pi_{lj} \) is the cumulative probability for each of the 12 categories at the stimulus contrast level \( x_i \), with \( x_1 = -2, x_2 = -1, x_3 = 0, x_4 = 1, \) and \( x_5 = 2 \); \( c_1 \cdots c_{11} \) are the cutpoints for 12 categorical responses. The scale parameter is exponentiated to ensure non-negative variance estimates. It can be shown that \( \exp(-\alpha) \) is the ratio of standard deviations between the reference distribution (zero contrast) and any of the distribution corresponding to the non-zero stimulus contrast.

Parameter Estimation by Non-Linear Regression

At present fitting procedures for models such as those specified by Eq. (1) do not exist in general purpose statistical software, although some procedures are available for specific cases (Kang Fu, 1998). Typically, specialized computer programs have been used. Tosteson and Begg used PLUM, an interactive package written in FORTRAN, for their analysis. On the other hand, standard non-linear regression programs are available in BMDP, SAS, and SPLUS. The approach proposed by Cox (1984) is to treat the problem of maximizing the likelihood as a regression problem.
(Jennrich & Moore, 1975). This requires a non-linear regression program that uses the Gauss-Newton algorithm in iteratively reweighted mode. The program used in this paper is SAS PROC NLIN (SAS Institute, Inc., 1999). Our presentation assumes only minimal knowledge of SAS and listings are provided for the first example. The price and availability of this statistical package can be obtained at www.sas.com.

To fit rating data by ordinal regression using PROC NLIN, the user specifies the predicted probabilities for each categorical response as a function of the parameters using Eq. (1) and the weights which are the inverses of the predicted frequencies multiplied by the sample size of the particular condition. The weights correspond to elements of a generalized inverse of the multinomial covariance matrix (see McCullagh & Nelder, 1989, p. 168). PROC NLIN iteratively updates the weights as the parameter estimates (and hence the predicted frequencies) are updated. By default, the residual sum of squares is used by PROC NLIN as the quantity to be minimized. For our purpose, the default loss function (LOSS is a SAS keyword) is replaced by the deviance. The deviance for the multinomial distribution is

\[ G^2 = -2 \sum_{ij} y_{ij} \ln(\hat{\pi}_{ij}/p_{ij}), \]

where \( y_{ij} \) is the observed count in category \( j \) of condition \( i \) and \( \hat{\pi}_{ij} \) is the fitted probability of \( j \)th categorical response for distribution \( i \). Differences between values of deviance at convergence can be used for comparing the goodness-of-fit of nested models. For large samples, the difference in \( G^2 \) is approximately distributed as \( \chi^2 \) with degrees of freedoms equal to the number of parameters of the full model minus the number of parameters of the reduced model. The Pearson \( \chi^2 \) statistic is a quadratic approximation for the \( G^2 \) statistic (Agresti, 1996). The higher order terms in the \( G^2 \) statistic becomes negligible only when the model holds and the sample size is sufficiently large. The \( G^2 \) statistic is typically adopted for model comparison in the generalized linear model framework.
The user can also specify the partial derivatives with respect to the model parameters or choose the derivative-free algorithm (which computes the required derivatives numerically) with the METHOD=DUD option. Other options in PROC NLIN allow the user to specify bounds on the values of the parameters, set a grid of possible starting values, and a criterion of convergence and a method of selecting the step-size used by the optimization algorithm.

The next section presents the results of fitting Eq. (1) to the rating data reported in Kijewski et al. (1989). The non-linear regression approach has two advantages: (1) The implementation does not require derivations of second-order derivatives. (2) The parameters specified in Eq. (1) are regression coefficients and, hence, can be directly evaluated. For instance, if the estimated value of the scale parameter $\alpha$ in our 13-parameter model is close to zero this suggests that an equal variance model can account for Kijewski et al.’s data.

Analysis of Rating Data from Multiple-Alternative Tasks

Listing 1 presents the SAS code for fitting Eq. (1) to multiple-alternative rating data with a continuous covariate representing the level of stimulus contrasts. The top and bottom seven lines in the input (data) file to the program are shown in Listing 2. The parameter estimates and the goodness-of-fit index (deviance) are shown in Listing 3. The value next to Objective in the output is the minimum deviance.

-----------------------------------------------------

Insert Listings 1, 2, and 3 about here

-----------------------------------------------------

Inspection of Listing 3 suggests an equal variance model applies because the estimate of the scale parameter, $\alpha$ (a in the SAS listing), is close to zero. This is suggests a 12-parameter model with one location parameter and 11 cutpoint param-
eters. This model corresponds to the equal-variance case of a normal model and is a member of the generalized linear model family. Thus, SAS PROC LOGISTIC with a probit link can also be used to fit the simpler 12-parameter model directly. The unequal-variance model as specified in Eq. (1), on the other hand, is not a member of the generalized linear models because it is nonlinear in the parameters. It should be pointed out that nonlinear in parameters does not imply nonlinear relationship between the z-transformed hits and false alarms. What it does imply is that the slope parameter for relating z-transformed false alarms to z-transformed hits is not equal to unity. The equal- and unequal-variance models are however nested and the difference between values of the $G^2$ statistic (deviance) can be used to test if the fits between the full and reduced models are significantly different. The results of the equal-variance model fit using SAS PROC NLIN, are shown in Listing 4. The deviance measures (Objective) are almost identical for both the full model and reduced model, indicating that there is no evidence for unequal variance. The $G^2$ value of 36.91 with 43 degrees of freedom indicates that the equal-variance normal models using the level of stimulus contrast as a quantitative covariate fits as well as Kijewski et al.'s 19-parameter model. They reported a $\chi^2$ value of 31.61 with 36 degrees of freedom.

---

Insert Listings 4 and 5 about here

---

Listing 5 displays the results of fitting the equal-variance normal model using SAS PROC LOGISTIC. As expected the results are the same as those obtained by the non-linear regression procedure, demonstrating that maximum-likelihood estimates of signal detection parameters can be obtained by non-linear regression
procedure, as discussed in the previous section. The intercept parameter estimates in the PROC LOGISTIC are the cutpoint estimates; the slope estimate (b in Listing 5) corresponds to the location-shift caused by a unit increase in the stimulus contrast level.

Application: Frequency Effect in Recognition Memory

McAuley and Heathcote (1999) conducted an experiment on recognition memory for words using a study-test paradigm. They used a within-subject design with three levels of word frequency (low, medium, and high) crossed with two levels of new (unstudied) versus old (studied) words. Thus, there were 6 treatment combinations. Subjects were asked to provide ratings from 1 to 6 to indicate their confidence that a word was new (1 = sure new to 3 = possibly new) or old (4 = possibly old to 6 = sure old). It was expected that subjects would be able to discriminate between new and old words and the discriminability would be highest for the condition using low frequency words and the lowest for the condition using high frequency words. To use the regression approach, we construct a design matrix for the two by three factorial design. We used one indicator variable to code the new-old word condition (Old = 1, New = 0), and two indicator variables for the three levels of word-frequency factor. For the word frequency factor, we used reference cell coding indicating membership of the medium and low frequency conditions respectively. The new-high word frequency condition serves as the reference group (intercept) in this regression model. Listing 6 shows the top 10 rows of the input file for a participant in the experiment.

-----------------------------

Insert Listing 6 about here

-----------------------------
Let $p_{i,j}$ be the probability of making a $j$th categorical response in the $i$th condition. A condition is defined by three indicator variables, $x_{i,1}$, $x_{i,2}$, and $x_{i,3}$, the first for new-old word condition, and the latter two for medium and low frequency words, respectively. We model the cumulative probability $\pi_{i,j} = p_{i,1} + \cdots + p_{i,j} \ (1 \leq j \leq k - 1, \ k = 6)$ by

$$
\pi_{i,j} = \Phi\left(\frac{c_1 + d_1 + \cdots + d_{j-1} - (\beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3})}{\exp(\alpha x_{i,1})}\right),
$$

where $c_1$ is the first cut point, and $d_l = c_{l+1} - c_l$, $l = 1 \cdots 4$. The distribution for latent responses evoked by the presentation of stimulus in the new-high frequency word condition assumes a standard normal distribution. In this formulation the ratio of standard deviations between the reference distribution and any of the signal condition is represented by $\exp(-\alpha)$. If $\alpha$ is not equal to zero, we have an unequal-variance model (Ratcliff, Sheu, & Gronlund, 1992). Preliminary model fitting achieved satisfactory goodness-of-fit of the model holding the value of the $\alpha$ parameter at a non-zero constant. This means that all the signal conditions relative to the reference condition have the same variance ratio. This is a much simpler model than one with a different scale parameter for each of the conditions relative to the reference condition. The parameter $\beta_1$ represents the location shift from the reference distribution to that of old-high frequency word condition; the sum $\beta_1 + \beta_2$ represents the location shift from the reference distribution to the of old-medium frequency word condition; and $\beta_1 + \beta_3$ represents the location shift from the reference distribution to the of old-low frequency word condition.

To illustrate the flexibility of the regression approach we test the hypothesis that location increases as word frequency decreases. To do so, we estimate the $\beta$ parameters under both the restricted-order constraint ($\beta_1 \geq 0$; $\beta_3 \geq \beta_2 \geq 0$) and the unrestricted condition ($\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_3 \geq 0$). To constrain order, we use a difference parameterization and bounding below by zero (The BOUNDS statement in SAS
restrains the parameter estimates within specific bounds. See, for example, Listing 1). Note that the parameter estimates cannot be constrained using the traditional fitting procedures of ROC data. Hence, one must conduct a test of homogeneity of means among the "population" of \( \beta \) parameters under restrictive alternatives such as those given above (Bartholomew, 1959) after extracting the parameter estimates. Unfortunately, the computation for such a test procedure is quite involved and is not available in standard statistical software packages. With our methods using the difference parameterization and zero bounding constraints, in contrast, the parameters can be estimated easily.

Our results show that the model fits data for 27 out of 32 subjects (not rejected at .05 level of significance) when the \( \beta \) parameters were not restricted. Inspection of the standardized residual plots reveals that the model fits were inadequate for only 4 subjects. The model estimation under order-constraint fit 23 out of 32 subjects (with the same 23 subjects among the 27 subjects in the unrestricted case). The parameter estimates are very similar for all but 3 of the 4 subjects where only the order-unrestricted model fit well. The average parameter estimates of 23 subjects (under order-constraint) are \( \beta_1 = 1.874, \beta_2 = .137, \beta_3 = .326 \) for the location shifts and for the ratio of standard deviations, \( \exp(-\alpha) = .717 \). We interpret this as a direct evidence in support of the word frequency effect in recognition memory experiment and conclude that there is evidence for a word frequency effect on location, with location being ordered by word frequency for the majority of subjects.

Conclusion

Signal detection models play an important role in the analysis of rating data from cognitive, perceptual, and memory experiments. The ordinal regression modeling methods proposed by Tosteson and Begg (1988) allow the experimenter to assess
the effects of treatment factors directly by examining regression coefficients and they allow for the possibility of both covariates adjusted location and scale parameters. Cox (1984) showed that the extensions of generalized linear models such as the general model proposed by Tosteson and Begg (1988) can be fitted by iteratively reweighted least squares. The use of SAS PROC NLIN provides a convenient and flexible platform for model fitting. The user can place bounds on the parameters as well as conduct an initial grid search. The derivative-free algorithm can be used when the derivatives are difficult to specify or the derivatives can be obtained from software that does symbolic computation. The advantage of the regression approach is that one does not need to devote a large amount of effort to modifying a specialized computer program in order to compare a variety of models, as is commonly required for nested model testing.


Acknowledgements

This work is supported in part by a paid-leave program from the University Research Council of DePaul University to Ching-fan Sheu and partly by a research visitor grant from the University of Newcastle, New South Wales, Australia.

Correspondence concerning this article should be addressed to Ching-Fan Sheu, Department of Psychology, DePaul University, 2219 North Kenmore Ave, Chicago, Illinois 60614-3522 (e-mail:csheu@depaul.edu).
Listing 1

Listing 1: SAS code to fit the normal model for multiple-alternative rating data in terms of Eq (1) using PROC NLIN. The location parameter $\beta$ is b and the scale parameter $\alpha$ is a in the SAS code.

/* Program starts:
* id: observation ID
* fr: frequency of a response category
* nt: total number of responses in a condition
* rc: rating categories
* x: indicator variable for contrast levels
* nc: number of response categories;
* rp = the observed response probability
*/

DATA a;
  INFILE 'ksj.asc';
  INPUT id nr nt categ x;
  rp = nr/nt;
  nc = 12;
RUN;

PROC NLIN METHOD=DUD;
  PARMS c1 = -3.5 d1 = .5 d2 = .5 d3 = .5 d4 = .5 d5 = .5
         d6 = .5 d7 = .5 d8 = .5 d9 = .5 d10 = .5 b = 1.5 a = .05;
  /* set distances between cutoff values to be positive */
  BOUNDS d1 > 0, d2 > 0, d3 > 0, d4 > 0, d5 > 0,
       d6 > 0, d7 > 0, d8 > 0, d9 > 0, d10 > 0;
ax = exp(a*x);

/* model response prob. of category one */

IF (categ = 1) THEN DO;
   zj = (c1 - b*x)*ax;
   MODEL rp = probnorm(zj);
END;

/* probnorm returns cumulative normal prob. */

/* model response prob. of categories 2 to next to last:
   express cutoffs by initial cutoff and the distances
   between successive cutoffs */

IF (categ > 1 AND categ < nc) THEN DO;
   IF (categ = 2) THEN DO;
      kj = c1 + d1;
      ko = c1;
   END;
   IF (categ = 3) THEN DO;
      kj = c1 + d1 + d2;
      ko = c1 + d1;
   END;
   IF (categ = 4) THEN DO;
      kj = c1 + d1 + d2 + d3;
      ko = c1 + d1 + d2;
   END;
   IF (categ = 5) THEN DO;
      kj = c1 + d1 + d2 + d3 + d4;
   END;
END;
ko = c1 + d1 + d2 + d3;
END;

IF (categ = 6) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5;
ko = c1 + d1 + d2 + d3 + d4;
END;

IF (categ = 7) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5 + d6;
ko = c1 + d1 + d2 + d3 + d4 + d5;
END;

IF (categ = 8) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7;
ko = c1 + d1 + d2 + d3 + d4 + d5 + d6;
END;

IF (categ = 9) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8;
ko = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7;
END;

IF (categ = 10) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8 + d9;
ko = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8;
END;

IF (categ = 11) THEN D0;
kj = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8 + d9 + d10;
ko = c1 + d1 + d2 + d3 + d4 + d5 + d6 + d7 + d8 + d9;
END;
zj = (kj - b*x)*ax;
zo = (ko - b*x)*ax;
pj = probnorm(zj);
po = probnorm(zo);
MODEL rp = pj - po;
END;

/* model prob. of the last response category */
IF (categ = nc) THEN D0;
    zj = (c1+d1+d2+d3+d4+d5+d6+d7+d8+d9+d10-(b*x))*ax;
    MODEL rp = 1 - probnorm(zj);
END;

/* generalized inverse of multinomial covariances */
_WEIGHT_ = nt/model_rp;
/* deviance is 2(n_i)log(p) - 2(n)log(p estimated) */
dev = -2*nr*log(model_rp);
IF (rp > 0 and rp < 1) THEN dev = dev + 2*nr*log(rp);
/* replace proc nlin default weighted loss function by deviance */
_LOSS_ = dev/_WEIGHT_;
pr = model_rp;
RUN;
/* Program ends */
Listing 2

Listing 2: The first and last 5 lines of data file to SAS program of Listing 1. The contrast level is treated as a quantitative variable. Column 1: Observation ID, Column 2: Number of responses, Column 3: Total number of responses in a condition, Column 4: Response category, Column 5: Stimulus contrast level.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>38</td>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>38</td>
<td>3</td>
<td>-2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>38</td>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>38</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>56</td>
<td>1</td>
<td>38</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>57</td>
<td>4</td>
<td>38</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>58</td>
<td>12</td>
<td>38</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>59</td>
<td>8</td>
<td>38</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>60</td>
<td>13</td>
<td>38</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>
Listing 3

Listing 3: Edited SAS output of PROC NLIN for unequal-variance normal models with the level of contrasts treated as a quantitative variable. The objective function to be minimized is the deviance.

The NLIN Procedure

Estimation Summary

<table>
<thead>
<tr>
<th>Objective</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>36.90741</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Approximate</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>-3.3685</td>
<td>0.2552</td>
<td>-3.8819</td>
<td>-2.8551</td>
</tr>
<tr>
<td>d1</td>
<td>0.8743</td>
<td>0.1928</td>
<td>0.4865</td>
<td>1.2620</td>
</tr>
<tr>
<td>d2</td>
<td>0.7006</td>
<td>0.1649</td>
<td>0.3689</td>
<td>1.0324</td>
</tr>
<tr>
<td>d3</td>
<td>0.5223</td>
<td>0.1438</td>
<td>0.2330</td>
<td>0.8115</td>
</tr>
<tr>
<td>d4</td>
<td>1.2979</td>
<td>0.2099</td>
<td>0.8756</td>
<td>1.7202</td>
</tr>
<tr>
<td>d5</td>
<td>0.5685</td>
<td>0.1465</td>
<td>0.2739</td>
<td>0.8632</td>
</tr>
<tr>
<td>d6</td>
<td>0.6116</td>
<td>0.1511</td>
<td>0.3077</td>
<td>0.9154</td>
</tr>
<tr>
<td>d7</td>
<td>0.5495</td>
<td>0.1445</td>
<td>0.2588</td>
<td>0.8402</td>
</tr>
<tr>
<td>d8</td>
<td>0.6723</td>
<td>0.1595</td>
<td>0.3515</td>
<td>0.9932</td>
</tr>
<tr>
<td>d9</td>
<td>1.0107</td>
<td>0.2044</td>
<td>0.5996</td>
<td>1.4219</td>
</tr>
<tr>
<td>d10</td>
<td>0.4888</td>
<td>0.1681</td>
<td>0.1507</td>
<td>0.8269</td>
</tr>
<tr>
<td>b</td>
<td>1.7070</td>
<td>0.1151</td>
<td>1.4755</td>
<td>1.9386</td>
</tr>
<tr>
<td>a</td>
<td>-0.00078</td>
<td>0.0632</td>
<td>-0.1279</td>
<td>0.1263</td>
</tr>
</tbody>
</table>
Listing 4

Listing 4: Edited SAS output of PROC NLIN for the equal-variance normal models. The model assumes the same scale parameter for the distributions of all conditions.

The NLIN Procedure

<table>
<thead>
<tr>
<th>Method</th>
<th>DUD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iterations</td>
<td>93</td>
</tr>
<tr>
<td>Objective</td>
<td>36.90754</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Approximate</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>-3.3680</td>
<td>0.2495</td>
<td>-3.8695</td>
<td>-2.8664</td>
</tr>
<tr>
<td>d1</td>
<td>0.8734</td>
<td>0.1814</td>
<td>0.5086</td>
<td>1.2381</td>
</tr>
<tr>
<td>d2</td>
<td>0.7004</td>
<td>0.1612</td>
<td>0.3763</td>
<td>1.0245</td>
</tr>
<tr>
<td>d3</td>
<td>0.5221</td>
<td>0.1417</td>
<td>0.2372</td>
<td>0.8071</td>
</tr>
<tr>
<td>d4</td>
<td>1.2980</td>
<td>0.2076</td>
<td>0.8806</td>
<td>1.7154</td>
</tr>
<tr>
<td>d5</td>
<td>0.5686</td>
<td>0.1448</td>
<td>0.2774</td>
<td>0.8597</td>
</tr>
<tr>
<td>d6</td>
<td>0.6116</td>
<td>0.1493</td>
<td>0.3114</td>
<td>0.9118</td>
</tr>
<tr>
<td>d7</td>
<td>0.5496</td>
<td>0.1424</td>
<td>0.2632</td>
<td>0.8360</td>
</tr>
<tr>
<td>d8</td>
<td>0.6728</td>
<td>0.1559</td>
<td>0.3593</td>
<td>0.9863</td>
</tr>
<tr>
<td>d9</td>
<td>1.0115</td>
<td>0.1910</td>
<td>0.6275</td>
<td>1.3955</td>
</tr>
<tr>
<td>d10</td>
<td>0.4896</td>
<td>0.1598</td>
<td>0.1683</td>
<td>0.8108</td>
</tr>
<tr>
<td>b</td>
<td>1.7070</td>
<td>0.1137</td>
<td>1.4783</td>
<td>1.9357</td>
</tr>
</tbody>
</table>
Listing 5

Listing 5: Edited SAS output of equal-variance normal model using SAS PROC LOGISTIC.

The LOGISTIC Procedure

Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>43</td>
<td>36.9075</td>
<td>0.8583</td>
<td>0.7317</td>
</tr>
<tr>
<td>Pearson</td>
<td>43</td>
<td>49.9615</td>
<td>1.1619</td>
<td>0.2163</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1</td>
<td>-3.3680</td>
<td>0.2446</td>
<td>189.6588</td>
<td>0.0001</td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>-2.4946</td>
<td>0.2106</td>
<td>140.2986</td>
<td>0.0001</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
<td>-1.7943</td>
<td>0.1859</td>
<td>93.1675</td>
<td>0.0001</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
<td>-1.2721</td>
<td>0.1710</td>
<td>55.3388</td>
<td>0.0001</td>
</tr>
<tr>
<td>C5</td>
<td>1</td>
<td>0.0259</td>
<td>0.1493</td>
<td>0.0300</td>
<td>0.8625</td>
</tr>
<tr>
<td>C6</td>
<td>1</td>
<td>0.5944</td>
<td>0.1531</td>
<td>15.0705</td>
<td>0.0001</td>
</tr>
<tr>
<td>C7</td>
<td>1</td>
<td>1.2060</td>
<td>0.1656</td>
<td>53.0559</td>
<td>0.0001</td>
</tr>
<tr>
<td>C8</td>
<td>1</td>
<td>1.7557</td>
<td>0.1816</td>
<td>93.4516</td>
<td>0.0001</td>
</tr>
<tr>
<td>C9</td>
<td>1</td>
<td>2.4285</td>
<td>0.2055</td>
<td>139.6923</td>
<td>0.0001</td>
</tr>
<tr>
<td>C10</td>
<td>1</td>
<td>3.4400</td>
<td>0.2467</td>
<td>194.4439</td>
<td>0.0001</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>3.9295</td>
<td>0.2695</td>
<td>212.5574</td>
<td>0.0001</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>1.7070</td>
<td>0.1115</td>
<td>234.2915</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Listing 6

Listing 6: The first 10 lines of data file from a participant in the word recognition experiment (McAuley & Heathcote, 1999). Column 1: Subject ID; Column 2: Response frequency in each of 6 response categories; Column 3: Total response frequencies for a condition; Column 4: Response categories; Column 5: Word frequency condition, 1 = High, 2 = Medium, 3 = Low; Column 6: New-Old word condition, New = 0, Old = 1; Column 7: Medium frequency indicator variable, H = 0, M = 1, L = 0; Column 8: Low frequency indicator variable, H = 0, M = 0, L = 1.

\[
\begin{array}{cccccccc}
S1 & 69 & 150 & 1 & 1 & 0 & 0 & 0 \\
S1 & 41 & 150 & 2 & 1 & 0 & 0 & 0 \\
S1 & 19 & 150 & 3 & 1 & 0 & 0 & 0 \\
S1 & 16 & 150 & 4 & 1 & 0 & 0 & 0 \\
S1 & 4 & 150 & 5 & 1 & 0 & 0 & 0 \\
S1 & 1 & 150 & 6 & 1 & 0 & 0 & 0 \\
S1 & 46 & 150 & 1 & 2 & 0 & 1 & 0 \\
S1 & 52 & 150 & 2 & 2 & 0 & 1 & 0 \\
S1 & 30 & 150 & 3 & 2 & 0 & 1 & 0 \\
S1 & 14 & 150 & 4 & 2 & 0 & 1 & 0 \\
\end{array}
\]