

# **Parameter Estimation of Signal Detection Models: RscorePlus User's Manual**

RscorePlus Version 5.6.1

22 September 2010

Lewis O. Harvey, Jr.

Department of Psychology and Neuroscience

Campus Box 345

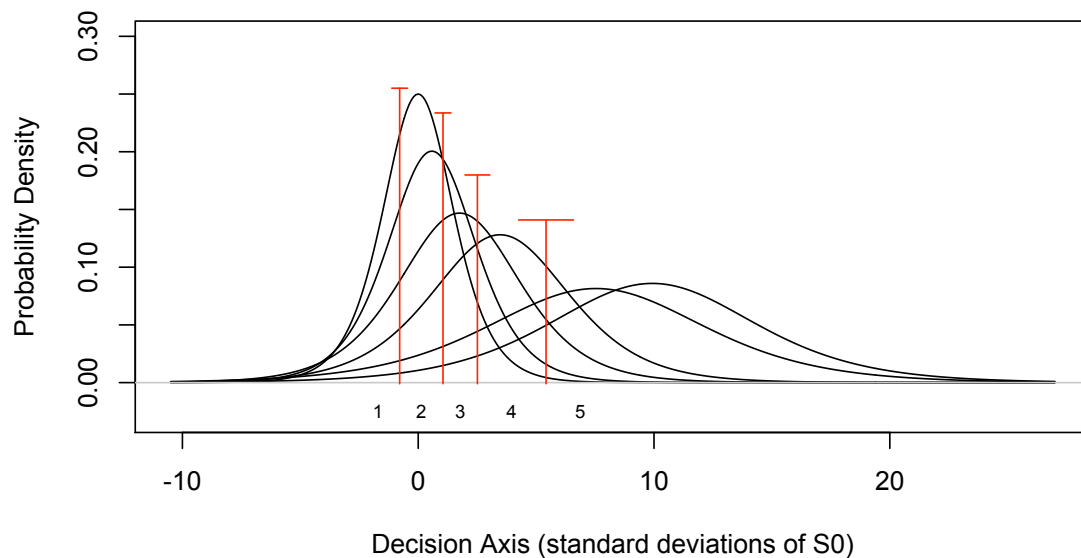
University of Colorado,

Boulder, Colorado 80309-0345 USA

<http://psych.colorado.edu/~lharvey/>

[lharvey@psych.colorado.edu](mailto:lharvey@psych.colorado.edu)

## **Signal Detection Model with Decision Criteria**



Blank Page

## **RscorePlus Signal Detection Rating-Scale Analysis**

### **Maximum-Likelihood Fit of Models to Data**

The present computer program, RscorePlus, has evolved from the FORTRAN program RSCORE written in the late 1960's by Donald Dorfman and Edward Alf, Jr., Department of Psychology, University of Iowa. That program was based on the method-of-scoring algorithm for fitting a nonlinear function to data (Dorfman & Alf, 1969; Dorfman, Beavers, & Saslow, 1973). In the intervening years more robust nonlinear fitting techniques have been developed to fit complex models to data. These newer methods are used in the RscorePlus. The original RSCORE could analyze data from two signal conditions (blank trials and signal trials, for example). The current RscorePlus application can handle up to 99 signal conditions for the single-interval paradigm. Three output files are produced: a Microsoft Word file suitable for printing, a text file for importing into a statistical analysis application (e.g., R) and a text file for importing into a graphics application (e.g. R or KaleidaGraph), from which one can produce a wide variety of graphs, as illustrated later.

RscorePlus is written in ANSI/ISO C/C++. It uses singular value decomposition (Press, Teukolsky, Vetterling, & Flannery, 2002), combined with a variation of the Marquardt method for nonlinear least-squares regression (Marquardt, 1963; Press, et al., 2002), to find the maximum-likelihood fit of the multiple-distribution, variable-criterion signal detection model to confidence rating-scale data. RscorePlus can fit models based on Gaussian, logistic, extreme value, gamma, and Poisson probability distributions. It can also fit the (mostly discredited) high-threshold model to the observed data in order to illustrate how poorly this model fits (Krantz, 1969). Data from four experimental paradigms may be analyzed: Single Interval,  $m$ -Alternative Forced-Choice, Same-Different and Match-To-Sample (ABX). The combinations of experimental paradigms and probability models that are available are shown in Table 1.

In a single-interval paradigm the observer is presented with one of  $m$  signal conditions on each trial and the observer must respond with one of  $n$  confidence ratings.

In the classical single-interval detection paradigm,  $m$  is equal to two: one signal condition represents blank trials and the other represents stimulus trials. For the other experimental paradigms (with the exception of the  $m$ -alternative, forced-choice paradigm with  $m$  greater than 2) the RscorePlus analysis treats all data sets as if they are single-interval data with two types of signal conditions. After the basic SDT model is fit, the detection parameters of the underlying signals are derived in a manner appropriate to the experimental paradigm. These relationships are found in Green and Swets (1966/1974) and in Macmillan and Creelman (2005).

**Table 1: Test Paradigms and Probability Distributions Handled by RscorePlus**

	Experimental Paradigms				
Probability Distributions	Single Interval	2AFC	mAFC > 2	Same-Different	ABX
Gaussian	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
Logistic	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
Extreme Minimum Value	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
Extreme Maximum Value	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
Gamma	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
Poisson	Confidence Ratings	Confidence Ratings	Interval Choices	Confidence Ratings	Confidence Ratings
High Threshold Model	Confidence Ratings				

### Assumptions of the Model

There are two broad classes of probability distributions that RscorePlus can fit to observed data: Class I distributions whose domain ranges from  $-\infty$  to  $+\infty$  (Gaussian, logistic, extreme minimum value and extreme maximum value distributions); and Class II distributions whose domain ranges from 0 to  $+\infty$  (gamma, Poisson distributions). In its native form each probability distribution is described by two parameters: a location parameter and a shape parameter. Each distribution may also be expressed in a

standardized form where the parameters are the mean and standard deviation. The use of the standardized form of a probability distribution facilitates comparison with the traditional Gaussian distribution.

In the Class I signal detection model each of the  $m$  signal conditions is represented by a Class I probability density function with mean  $\mu_j$  and standard deviation  $\sigma_j$  where  $j$  ranges from 0 to  $m-1$ . It is usually assumed that  $\mu_0 = 0$  and  $\sigma_0 = 1$ , although the program allows other values to be specified. It is further assumed that the observer using  $n$  response categories holds  $n-1$  decision criteria  $X_c$  along a single decision axis. The units of the decision axis are  $\sigma_0$ . This program finds the means  $\mu_j$  and standard deviations  $\sigma_j$  of the remaining signal distributions 1 through  $m-1$ , and the  $n-1$  decision criteria  $X_c$  along the horizontal axis relative to the first signal distribution. The decision criteria are also reported in six other metrics, computed relative to the first ( $j = 0$ ) signal condition: the likelihood ratio ( $\beta$ ), the natural logarithm of  $\beta$ , the response bias index,  $c_a$ , the natural logarithm of the *a posteriori* odds that a signal was present, the *a posteriori* probability of a signal at each decision criterion  $P(s | X_c)$ , the *a posteriori* probability of getting a hit using each of the decision criteria. Detection sensitivity is reported in three different forms:  $d_a$ ,  $A_z$ , and  $pc_{max}$  (Simpson & Fitter, 1973). Performance is reported in terms of the critical hit rate, the critical false alarm rate and critical decision criterion required to achieve posterior hit probabilities of 0.75 and 0.90 (Harvey, 1992).

The program makes an initial estimate of the parameters from the observed data using a least-squares linear regression method. These least-squares estimates are then used as the input to the Marquardt nonlinear least-squares procedure (Marquardt, 1963; Press, et al., 2002) to find the maximum-likelihood estimates of the parameters. If the procedure does not converge on a solution within 99 iterations, it will terminate and the current parameter values will be printed.

There are three options for dealing with response frequencies that are zero, where  $m$  is the number of response ratings:

1. Change each zero response frequency to  $(1/m)$ ;

2. Add  $(1/m)$  to each response frequency, the so-called log-linear correction (Hautus & Lee, 1998);
3. Collapse adjacent response categories in order to eliminate all zero frequencies in the data.

Previous versions of RscorePlus automatically eliminated response categories that had no responses at all. The current version retains these response categories, converting the zeros to small values according to the option selected. This strategy means that multiple sets of data using the same number of response categories will have the same number of decision criteria in the best-fitting model. Of course the two criteria that form the boundary of the category having no responses will be very close together (but not identical, as they should be if the response frequency were really zero).

You could also combine the sparse data of several subjects to eliminate zero frequencies. The costs and benefits of combining data from different subjects are discussed by Macmillan (Macmillan & Creelman, 2005; Macmillan & Kaplan, 1985). If observers used only two response categories (for example, *yes* and *no*), the program assumes the equal-variance model ( $\sigma_n = \sigma_s = 1$ ). At the user's option, another value of the signal standard deviation may be provided. It will be held constant and only the signal means will be fitted.

### Input Data

This version of the program will handle 99 confidence-rating levels and 99 signal conditions. Note that since version 5.5 of RscorePlus there is an additional input parameter in the data file (`paradigmType`). Data files used with previous versions will have to have this parameter added. There is also a single line that must be read in before any of the data sets. This single line at the beginning of the file is the heading that will be used in the alternate output file. If you put tab separated items in the title of each data set, you can define what the columns mean in this first line by having tab-separated column headers in this first line. When you read the alternate output file into R or other applications, the columns will line up with the appropriate header label. See `test2.txt` sample data file for an example of how to use the tab-separated labels. Although the data

in the input text file are read in as a stream it is more convenient to arrange the input data like this:

Header line: Tab-separated header labels for all the data sets as defined in their individual titles

Input Line 1: Alphanumeric Title String (maximum of 255 characters)

Input Line 2:

NumberOfRatings:	Number of rating categories
NumberOfSignals:	Number of signal conditions
SubjectIDNumber:	Subject number
CollapseFlag:	Flag for collapsing after unsuccessful analysis (see below)
NoZerosFlag:	Flag for eliminating all zeros from data (see below)
AltOutFlag:	Flag for alternate output options (see below)
ParadigmType:	SINT, NAFC, SDIF, or ABX to indicate the paradigm (see below)

Input Line 3:

The word 'label' followed by 8-character labels for each of the *NumberOfRatings* response categories. No spaces are allowed in the labels. The labels may be shorter than eight characters.

Data Input Lines:

For each signal condition the input data consist of an 8-char label (maximum length) followed by the frequency of each rating category ordered by increasing confidence that the signal was present. Successive signal conditions should be ordered by increasing signal intensity. The program will stop when the value of *NumberOfRatings* read on input equals zero (writes a banner) or is negative (suppresses the banner).

Starting Parameter Values:

The starting values of the *NumberOfSignals* means and then the starting standard deviations. The classical model sets the mean of the first condition at 0.0 and its standard deviation at 1.0. You may use other values as appropriate: they will be held constant during the fitting process. The other means are free parameters. The other standard deviations are free parameters if the given value is 0.0; otherwise the supplied value will be used and held fixed.

## Options Available:

- fCollapse      = 0 Do not collapse data after unsuccessful analysis  
                 = 1 Collapse data after unsuccessful analysis
- fNoZeros       = 0 Use (1 / Number of Ratings) in place of zero frequencies  
                 = 1 Use log-linear correction to eliminate zeros  
                 = 2 Collapse categories to eliminate zeros
- fAltOut        = 0 Do not write parameters to alternate files  
                 = 1 Write solution parameters to alternate files
- ParadigmType= SINT (single interval experimental paradigm)  
                 = MAFC (m-alternative, forced-choice experimental paradigm)  
                 = SDIF (same-different experimental paradigm)  
                 = ABX (match-to-sample or ABX paradigm)



## Sample Input

Here is a set of data with two signal conditions (signal vs. blank) taken from Swets, Tanner, and Birdsall (1961) and included in data set test1.txt:

```

Headings
TEST SUBJECT NUMBER 1 FROM SWETS, TANNER, AND BIRDSALL
6 2 1 0 0 1 SINT
labels R_1 R_2 R_3 R_4 R_5 R_6
blank 174 172 104 92 41 8
stimulus 46 57 66 101 154 173
0.0 1.0 0.5 1.0
0 0 1 1
end of file
0
Here is a set of data (test2.txt) with five stimulus intensities and a blank trial condition:
Subject Eye Visual Field Retinal Locus Distance Date
EP Left Eye Left VF 5 DEG D=22 CM 20.4.82
5 6 1 0 0 1 SINT
label rat_1 rat_2 rat_3 rat_4 rat_5
blank 13 16 9 2 0
2.5 9 10 9 5 0
2.4 6 7 6 10 3
2.3 3 6 3 13 9
2.2 0 2 2 2 17
2.1 0 1 1 2 24
0.0 1.0 0.5 1.0 1.0 1.0 1.5 1.0 2.0 1.0 2.5
1.0
0 0 1 1 1 1 1 1 1 1 1
1
end of file
-1
Here are five sets of data (test3.txt) illustrating the five types of experimental paradigms
handled by RscorePlus, taken from MacMillan and Creelman (1991, 2005):
Headings
Yes-No Example: Macmillan & Creelman, p 126
2 2 1 0 0 1 SINT
labels No Yes
blank 96 4
signal 60 40
0.0 1.0 1.0 1.0
0 0 1 0
2AFC Example: Macmillan & Creelman, p 120
2 2 2 0 0 1 NAFC 2
labels Rat_1 Rat_2
NewOld 18 7
OldNew 9 16
0.0 1.0 1.0 1.0
0 0 1 0
3AFC Example: Macmillan & Creelman, p 140
3 3 3 0 0 1 NAFC 3
labels Int_1 Int_2 Int_3
s1s0s0 4 3 3
s0s1s0 2 6 2
s0s0s1 1 1 8
0.0 1.0 1.0 1.0 1.0
0 0 1 1 1
Same-Different Example: Macmillan & Creelman, p 142
2 2 4 0 0 1 SDIF
labels Same Diff
Same 40 10
Diff 20 30
0.0 1.0 1.0 1.0
0 0 1 1
ABX Example: Macmillan & Creelman, p 164
2 2 5 0 0 1 ABX
labels B A
X=B 40 10
X=A 20 30
0.0 1.0 1.0 1.0
0 0 1 1
end of file
0

```

## Printed Output

For each set of data the program generates two pages of printed output in a text file with the name of the input file with “\_out.txt” appended. On Macintosh computers the creator of this file is MSWD so that Microsoft Word will start if you double click on the file. Each output page will fit on a Word page if you set the font to Courier, and the font size to 8. The output contains the following information:

### Printed Output: Page 1

The estimates of the signal detection parameters are based on the maximum-likelihood solution, where possible, or the least-squares solution where the maximum-likelihood solution is not possible. The type of solution is indicated in the heading. The following information is printed.

**Output Header:** Provided by the computer program.

**Data Identification:** Information taken from your input data

**Free Parameters in Model:** The number of fitted parameters in the model.

**Log Likelihood of Model:** The natural logarithm of the likelihood that the observed data have been generated by the best-fitting model. The fitting algorithm finds the values of the model parameters that maximizes this value. There is a simple relationship between the log likelihood and both chi-square and the AIC (see below).

**Chi-Square:** The chi-square statistic,  $\chi^2$ , computed from the log likelihood of the fit, is used as a goodness-of-fit measure (see Chapter 17 in Hays, 1973).  $\chi^2 = -2 \ln(\text{likelihood})$ .

If the final value of  $\chi^2$  is significant for the degrees of freedom indicated, it means that the detection model does **not** account for all the variance in the data and probably should be rejected. If  $\chi^2$  is not statistically significant, we cannot reject the hypothesis that the model provides a satisfactory description of the behavior in the rating scale detection experiment. In other words, a non-significant, low  $\chi^2$  means a good fit between the model and the data. The fitting algorithm finds a maximum-likelihood solution by minimizing  $\chi^2$  (which is equivalent to maximizing the likelihood). If there are zero

degrees of freedom, as when only two response categories exist,  $\chi^2$  is zero because the fit will always be perfect.

**AIC:** The Akaike Information Criterion (Akaike, 1974). AIC is a measure that allows comparison of the fit of different models having different number of free parameters. It is computed from the log likelihood with an adjustment for the number of free parameters:

$$\chi^2 = -2 \ln(\text{likelihood}) + 2 (\text{number of free parameters})$$

The smaller the AIC value, the better the fit of the model to the data.

**Model Parameters:** The means and standard deviations of the best-fitting probability distributions for each of the signal conditions. For non-Gaussian probability distributions the appropriate parameters will be given (i.e., location and shape parameters or mean and standard deviation). The standard errors of these parameters are also given.

**Decision Criteria:** The maximum-likelihood values of the  $n$ -1 decision criteria  $X_c$  are printed along with their standard errors. These criteria are expressed in standard deviation or shape parameter units (whichever is appropriate) of the first signal distribution, with zero on the scale corresponding to the mean or location parameter (whichever is appropriate) of that distribution.

**Detection Indices for Each Pair of Signal Conditions:** Three measures of detection sensitivity are computed for each pair of signal conditions and their standard errors. They are  $d_a$ ,  $A_z$ , and  $pc_{\max}$ . The first,  $d_a$ , is appropriate for the unequal-variance case and is numerically equal to  $d'$  in the equal variance case. The second,  $A_z$ , corresponds to the area under the ROC and is useful for expressing sensitivity in terms of a probability, since it is the percent correct that would be achieved in the 2-alternative, forced-choice paradigm. The third index,  $pc_{\max}$ , is the percent correct that would have been obtained if the decision criterion had been unbiased.

$$d_a = \frac{(\mu_j - \mu_i)}{\sqrt{\frac{\sigma_j^2 + \sigma_i^2}{2}}}$$

$$A_z = z^{-1} \left[ \frac{d_a}{\sqrt{2}} \right]$$

$$pc_{\max} = z^{-1} \left[ \frac{d_a}{2} \right]$$

**Single-Interval Equivalent Discrimination Indices:** These indices are only printed if the data are from one of the experimental paradigms other than single-interval. In these cases there are only two signals and these indices,  $d_a$ ,  $A_z$ , and  $pc_{\max}$ , are what would be achieved if the two underlying signals were used in a single-interval experiment. They are computed from the single-interval  $d_a$  representing the discriminability of the two types of trials (e.g., same and different) using equations from Chapter 5 (Forced-Choice Designs), Chapter 6 (Same-Different Designs), and Chapter 7 (ABX Designs) of Macmillan and Creelman (2005).

$$\text{2AFC} \quad d_{a,2afc} = \frac{d_{a,si}}{\sqrt{2}} \quad \text{Equations 5.2, 5.12}$$

$$\text{Same-Different} \quad d_{a,same-different} = 2 \cdot z \left[ 0.5 \left( 1 + \sqrt{2 \cdot pc_{\max,si} - 2} \right) \right] \quad \text{Equation 6.3}$$

$$\text{ABX} \quad pc_{\max,si} = z^{-1} \left[ \frac{d_{a,abx}}{\sqrt{2}} \right] z^{-1} \left[ \frac{d_{a,abx}}{2} \right] + z^{-1} \left[ \frac{-d_{a,abx}}{\sqrt{2}} \right] z^{-1} \left[ \frac{-d_{a,abx}}{2} \right] \quad \text{Equation 7.2}$$

**Single-Interval Equivalent Discrimination Indices for mAFC:** When the number of alternative forced-choice intervals is greater than 2 the data can not be analyzed by using the single-interval strategy employed with the other paradigms. In this case the input data are treated as a confusion matrix, with the response categories corresponding to the number of times that each interval was selected. The probability of being correct is computed from the major diagonal and this percent correct is used to compute the equivalent  $d_a$  by solving the inverse of Equation 3.10, p. 69, in Green and Swets (1966/1974) under the assumption of  $\mu = 0$  and  $\sigma = 1$ :

$$P_m(C) = \int_{-\infty}^{\infty} \left[ \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left(\frac{x-d_a-\mu}{\sigma}\right)^2} \left( \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2} \right)^{m-1} \right] dx$$

**Critical Performance Relative to the First Signal Condition:** The critical hit rate, critical false alarm rate, and the critical decision criterion required to achieve a posterior probability of a hit is given for posterior probabilities of 0.75 (“Clear and convincing evidence”) and 0.90 (“Beyond reasonable doubt”). These values apply to the prior signal

probabilities observed in the data. They can only be compared to those from other sets of data if the prior probabilities are the same. A second set of critical hit rates is also printed under the assumption that the priori signal probabilities were equal (e.g., 0.5). These values may be compared with each other across conditions and data sets. A fuller description of these measures is found in Harvey (1992).

**Response Bias c-sub-a:** A measure of response bias at each of the decision criterion points (Equation 3.13, page 74 of Macmillan & Creelman, 1991):

$$c_a = -\frac{\sigma_j}{\sigma_0 + \sigma_j} \sqrt{\frac{2}{\sigma_0^2 + \sigma_j^2}} [z(HR) - z(FAR)]$$

This index is weighted by the root mean square of the standard deviations of both the signal and the noise distributions and is appropriate when the fitted signal detection model has unequal variances.  $C_a$  is equal to bias measure  $C$  (Equation 2.2, page 33 of Macmillan & Creelman, 1991) when the model probability distributions have equal variances:

$$c = -\frac{\sigma_j}{\sigma_0 + \sigma_j} [z(HR) - z(FAR)] = -0.5 [z(HR) - z(FAR)]$$

**Beta:** The decision criteria expressed as likelihood ratios. Beta is given by:

$$\beta = \frac{f_j(X_c)}{f_1(X_c)}$$

where  $X_c$  is the criterion value described above and  $f(X_c)$  is the probability density function with mean ( $\mu$ ) and standard deviation ( $\sigma$ ) as appropriate for the noise and signal plus noise distributions. The Gaussian distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2}$$

**Ln-B:** The decision criteria expressed as the natural logarithm of the likelihood ratio  $\beta$ . For the purposes of statistical analyses it is better to use  $X_c$  or  $\ln\beta$  than  $\beta$  because  $\beta$  has a very skewed distribution.

**Ln-O:** The decision criteria expressed as the natural logarithm of the *a posteriori* odds favoring the signal at each of the  $X_c$  decision criterion points. These odds are based on  $\beta$  taking the *a priori* stimulus probabilities into account.

**P(signal|X):** The decision criteria expressed as the *a posteriori* probability of a signal given an x-axis observation placed at the corresponding criterion point. P(signal|X) takes both the *a priori* stimulus probabilities and the decision criterion into account.

**P(hit|X):** The *a posteriori* probability of a hit using the corresponding decision criterion,  $X_c$ , to respond yes or no. P(hit|X) takes into account the *a priori* stimulus probabilities and the decision criterion.

**SE (Standard Errors of the Parameters):** The value in parentheses below each parameter is the standard errors of the estimate of that parameter:  $\mu_j$ ,  $\sigma_j$ ,  $d_a$ ,  $A_z$ ,  $pc_{\max}$ , and  $X_c$ . They may be used for statistical hypothesis testing in the same manner that the standard error of the mean is used. The 95 percent confidence interval for a parameter value is the value  $\pm 1.96$  standard error.

## Printed Output: Page 2

**Output Header:** Provided by the computer program.

**Data Identification:** Information taken from your input data

**Observed response frequencies:** The observed raw frequencies of response for each of the  $n$  response categories for each of the  $m$  signal condition. The right-most number is the sum of the frequencies. These are your raw data so check them to make sure that you have made no errors in entering them. The probabilities below are based on the raw data after zero frequencies have been eliminated by one of the three options.

**Observed Response Probabilities:** The response frequencies each divided by the total number of responses in the corresponding signal condition. The right-most number is the *a priori* probability of that stimulus condition.

**Observed A Posteriori Probabilities:** The posterior probabilities of each signal condition  $s$ , given response  $R$ :  $P(s|R)$ .

**Observed Cumulative Probabilities:** The cumulative probabilities calculated by successively adding the above response probabilities starting with the highest criterion response category ("most certain that the signal was present"). A pair of cumulative

probabilities (one from one signal condition and one from another signal condition) may be interpreted as a HR–FAR pair. Blank trials generate false alarm rates (FAR) and signal trials generate hit rates (HR). The leftmost pair of cumulative probabilities is always 1, 1. Excluding this last pair, an  $n$ -category response scale therefore gives rise to  $n - 1$  HR–FAR pairs. These  $n - 1$  pairs may be plotted on linear coordinates to form the Receiver Operating Characteristic (ROC).

**Observed Cumulative Z-Score Probabilities:** The z-score transforms of the above cumulative probabilities based on the integral of the unit normal probability distribution (equivalent to the quantile function of the probability distribution with a mean of zero and a standard deviation of 1). Note that the sign of these numbers follows standard statistical convention: negative z-scores correspond to probabilities less than 0.5. These  $n - 1$  pairs may be plotted on linear z-score coordinates to form an ROC. The model predicts that the ROC will be a straight line.

**Predicted Response Frequencies:** The response frequencies predicted by the best-fitting model, corresponding to the observed response frequencies.

**Predicted Response Probabilities:** The response probabilities predicted by the best-fitting model, corresponding to the observed response probabilities.

**Predicted *A Posteriori* Probabilities:** The posterior probabilities,  $P(s|R)$ , predicted by the best-fitting model, corresponding to the observed posterior probabilities.

**Predicted Cumulative Probabilities:** The cumulative probabilities predicted by the best-fitting model, corresponding to the observed cumulative probabilities above. These points all lie on the predicted probability ROC.

**Predicted Cumulative Z-Score Probabilities:** The z-scores of the cumulative probabilities predicted by the best-fitting model. These points form the predicted linear z-score ROC.

The printed output from the first sample data set given above is reproduced below. You can use this output to make sure that your version of the program is running correctly.

Detection Model: Standardized Gaussian; Paradigm: SingleInterval  
 Maximum-likelihood fitting method by Lewis O. Harvey, Jr.  
 07:03:45, Wednesday, 20 Sep 2006 (Version 5.5.6)  
 Input data set 1 was read from file 'test1.txt'

Subject 1: TEST SUBJECT NUMBER 1 FROM SWETS, TANNER, AND BIRDSALL  
 There are 6 response categories and 2 signal conditions  
 No Auto-Collapse      Zeros->1/m      Results file open  
 There are 6 non-zero rating categories.

Maximum-likelihood solution reached in 6 iterations:  
 Goodness of fit: Standardized Gaussian, SingleInterval  
 Free Parameters in Model = 7, Log Likelihood of Model = -0.7409  
 Chi-Square(3) = 1.482, p = 0.68648, AIC = 15.482

....mean and sdev of the signal distributions.....

	mean	sdev	meanSE	sdevSE
blank:	0.0000	1.0000	-----	-----
stimulus:	1.5184	1.4165	0.0960	0.0631

....Decision criteria of the Standardized Gaussian model.....

	1 2	2 3	3 4	4 5	5 6
Xc:	-0.5331	0.2041	0.7098	1.3661	2.2926
se:	0.0536	0.0495	0.0526	0.0671	0.1133

....Detection indices of the signals relative to each other.....

	d(a)	A(z)	PCmax	da se	Az se	PC se
s0-s1:	1.2384	0.8094	0.7321	0.0662	0.0127	0.0109

....Critical performance relative to the first signal condition (s0)....

	prior	HR(75)	FAR(75)	Xc(75)	HR(90)	FAR(90)	Xc(90)
stimulus:	0.5025	0.7182	0.2418	0.7005	0.4725	0.0530	1.6162

....Critical performance relative to the first signal condition (s0)....

	prior	HR(75)	FAR(75)	Xc(75)	HR(90)	FAR(90)	Xc(90)
stimulus:	0.5000	0.7157	0.2386	0.7110	0.4705	0.0523	1.6231

....Response bias c-sub-a.....

	1 2	2 3	3 4	4 5	5 6
stimulus:	-0.9473	-0.3461	0.0665	0.6020	1.3575

....Likelihood ratio, beta.....

	1 2	2 3	3 4	4 5	5 6
stimulus:	0.2851	0.4687	0.7717	1.7845	8.4194

....Log likelihood ratio, log beta.....

	1 2	2 3	3 4	4 5	5 6
stimulus:	-1.2548	-0.7578	-0.2592	0.5792	2.1305

....Log odds in favor of signal.....

	1 2	2 3	3 4	4 5	5 6
stimulus:	-1.2447	-0.7477	-0.2491	0.5893	2.1406

....a posteriori probability of signal: P(s|Xc).....

	1 2	2 3	3 4	4 5	5 6
stimulus:	0.2236	0.3213	0.4381	0.6432	0.8948

....a posteriori probability of a hit: P(hit|Xc).....

	1 2	2 3	3 4	4 5	5 6
stimulus:	0.5710	0.6649	0.7517	0.8645	0.9643



Detection Model: Standardized Gaussian; Paradigm: SingleInterval  
 Maximum-likelihood fitting method by Lewis O. Harvey, Jr.  
 07:03:45, Wednesday, 20 Sep 2006 (Version 5.5.6)  
 Input data set 1 was read from file 'test1.txt'

Subject 1: TEST SUBJECT NUMBER 1 FROM SWETS, TANNER, AND BIRDSALL  
 There are 6 response categories and 2 signal conditions  
 No Auto-Collapse      Zeros->1/m      Results file open  
 There are 6 non-zero rating categories.

....Observed Response Frequencies.....

	R_1	R_2	R_3	R_4	R_5	R_6	totals
blank:	174.0	172.0	104.0	92.0	41.0	8.0	591.0
stimulus:	46.0	57.0	66.0	101.0	154.0	173.0	597.0
Totals:	220.0	229.0	170.0	193.0	195.0	181.0	1188.0

....Observed Response Probabilities.....

	R_1	R_2	R_3	R_4	R_5	R_6	prior
blank:	0.2944	0.2910	0.1760	0.1557	0.0694	0.0135	0.4975
stimulus:	0.0771	0.0955	0.1106	0.1692	0.2580	0.2898	0.5025
Prob:	0.1852	0.1928	0.1431	0.1625	0.1641	0.1524	1.0000

....Observed A Posteriori Probabilities.....

	R_1	R_2	R_3	R_4	R_5	R_6
blank R:	0.7909	0.7511	0.6118	0.4767	0.2103	0.0442
stimulus R:	0.2091	0.2489	0.3882	0.5233	0.7897	0.9558

....Observed Cumulative Probabilities.....

	1 2	2 3	3 4	4 5	5 6
blank:	1.0000	0.7056	0.4146	0.2386	0.0829
stimulus:	1.0000	0.9229	0.8275	0.7169	0.5477

....Standardized Quantile Score of Observed Cumulative Probabilities....

	1 2	2 3	3 4	4 5	5 6
blank:	+INFIN	0.5401	-0.2155	-0.7106	-1.3860
stimulus:	+INFIN	1.4254	0.9441	0.5733	0.1197

....Predicted Response Frequencies.....

	R_1	R_2	R_3	R_4	R_5	R_6	totals
blank:	175.5	167.8	106.5	90.4	44.3	6.5	591.0
stimulus:	44.0	61.5	64.1	103.4	149.5	174.5	597.0
Totals:	219.5	229.3	170.6	193.8	193.9	181.0	1188.0

....Predicted Response Probabilities.....

	R_1	R_2	R_3	R_4	R_5	R_6	prior
blank:	0.2970	0.2839	0.1802	0.1530	0.0750	0.0109	0.4975
stimulus:	0.0738	0.1030	0.1073	0.1731	0.2505	0.2923	0.5025
Prob:	0.1848	0.1930	0.1436	0.1631	0.1632	0.1523	1.0000

....Predicted A Posteriori Probabilities.....

	R_1	R_2	R_3	R_4	R_5	R_6
blank R:	0.7994	0.7318	0.6244	0.4665	0.2287	0.0357
stimulus R:	0.2006	0.2682	0.3756	0.5335	0.7713	0.9643

....Predicted Cumulative Probabilities.....

	1 2	2 3	3 4	4 5	5 6
blank:	1.0000	0.7030	0.4191	0.2389	0.0860
stimulus:	1.0000	0.9262	0.8232	0.7159	0.5428

....Standardized Quantile Score of Predicted Cumulative Probabilities...

	1 2	2 3	3 4	4 5	5 6
blank:	+INFIN	0.5327	-0.2037	-0.7095	-1.3663
stimulus:	+INFIN	1.4486	0.9277	0.5704	0.1072

### **Pooling of Response Categories**

If the program is unable to find the maximum-likelihood solution it will print the appropriate message. This condition can occur when one or more response categories have zero frequency of use in the raw data. Under these circumstances one strategy is to collapse two adjacent response categories (for both signal conditions) into a single category. If fCollapse flag is set to 1 the program will collapse the two categories with the lowest joint probabilities. It will then try again to find a maximum-likelihood solution. This process continues until either a solution is found or until only three categories remain. In the latter case, a least-squares solution is the only one possible (a straight line can be fit to two points without any error). Note: this auto-collapse option is disabled in the current version of RscorePlus.

A second approach is to set fNoZeros to 2 (see input flags) and the program will collapse all categories containing a zero frequency into an adjacent category before starting the analysis. The collapsing process continues until all zero frequencies are gone. Other strategies for dealing with zeros and how to collapse data across subjects are given by Macmillan and Kaplan (1985).

### **Abbreviated Output**

If the fAltOut flag is set to 1, the program will write three additional text files. The first file has the same name as the input file with '\_alt.txt' appended to the end. It contains tab-separated data of the model with a tab-separated heading. There is one row per data set. This text file can be directly imported by other applications such as Excel and R for further analysis. Note: The alternate output option should only be used with data sets having the same number of signal conditions and response categories because the number of data columns is determined by these numbers.

The second text file has the same name as the input file with '\_grf.txt' appended to the end. It contains a variety of observed and predicted data that are suitable for importing into a graphics program like KaleidaGraph or R for producing beautiful ROC curves, COC curves and other types of graphs as well. There is a tab-separated heading

that gives reasonably understandable labels for each of the data columns. An example of the heading for a two-signal analysis (signal 0 is a blank trial; signal 1 is a signal trial) is given below:

1. data\_set      The number of the data set in a multiple data set file
2. n\_ratings     The number of response categories
3. n\_signals     The number of signal conditions
4. Model        The type of probability distribution in the model
5. Scale        Whether the parameters are native or mean and standard deviation
6. Xc            The computed decision criterion points and a set of continuous values
7. Xc\_se        The estimated standard errors of the computed model Xc's.
8. Obs\_pYs0     Observed probability of a "yes" response given signal 0 (false alarm rate)
9. Obs\_pYs1     Observed probability of a "yes" response given signal 1 (hit rate)
10. Obs\_zYs0    Z-score of observed false alarm rate
11. Obs\_zYs1    Z-score of observed hit rate
12. Obs\_pHitXc0   Observed probability of a false alarm at each decision criterion
13. Obs\_pHitXc1   Observed probability of a hit at each decision criterion
14. Pred\_pYs0    Predicted probability of signal 0 given a yes response (false alarm rate)
15. Pred\_pYs1    Predicted probability of signal 1 given a yes response (hit rate)
16. Pred\_zYs0    Z-score of predicted false alarm rate
17. Pred\_zYs1    Z-score of predicted hit rate
18. Pred\_pHitXc0   Predicted probability of a false alarm at each decision criterion
19. Pred\_pHitXc1   Predicted probability of a hit at each decision criterion
20. Pred\_Density1   Predicted probability density function for signal 0
21. Pred\_Density2   Predicted probability density function for signal 1
22. Criteria      Vertical lines for each of the decision criteria
23. Rating\_Category   Integers 1 through NumberOfRatings
24. Rating\_Labels    If numeric (e.g., confidence probabilities), can be used for plotting
25. Obs\_Freq1     Observed response frequencies for blank trials
26. Obs\_Freq2     Observed response frequencies for signal trials
27. Pred\_Freq1    Predicted response frequencies for blank trials
28. Pred\_Freq2    Predicted response frequencies for signal trials
29. Obs\_Prob1     Observed response probabilities for blank trials
30. Obs\_Prob2     Observed response probabilities for signal trials
31. Pred\_Prob1    Predicted response probabilities for blank trials
32. Pred\_Prob2    Predicted response probabilities for signal trials
33. Obs\_pSr0      Observed posterior probability of blank given response R
34. Obs\_pSr1      Observed posterior probability of signal given response R
35. Pred\_pSr0     Predicted posterior probability of blank given response R
36. Pred\_pSr1     Predicted posterior probability of signal given response R

The third text file has the same name as the input file with '\_vcv.txt' appended to the end. It contains the variance-covariance matrix of the parameter values computed during the maximum-likelihood fitting procedure. The items are tab separated. Labels for the rows and columns are provided. The subject number for each data set is in position 1,1 of each matrix.

## Graphs

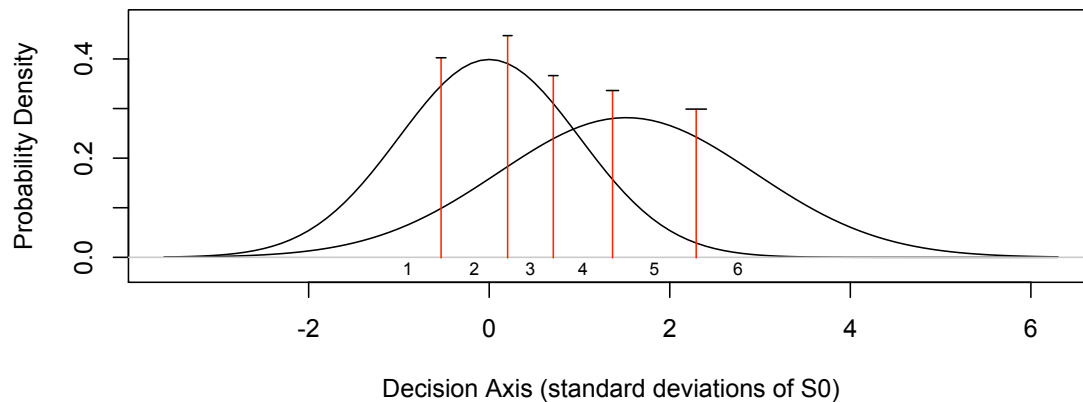
A wide variety of graphs may be produced by plotting various combinations of the variables in the \*\_grf.txt file. Some useful graphs based on the data in test1.txt (subject 1 from Swets, et al., 1961), prepared with R (version 3.8.1) on a Macintosh computer, are reproduced below. Other programs, such as KaleidaGraph, can be used, of course. The R code for producing these graphs is reproduced in Appendix IV and in the text file “SDT\_Graphs.R” in the download package.

### Signal Detection Model with Decision Criteria.

X-axis: “Xc” and standard error of Xc

Y-axis: “Pred\_Density0”, “Pred\_Density1” and “Criteria”

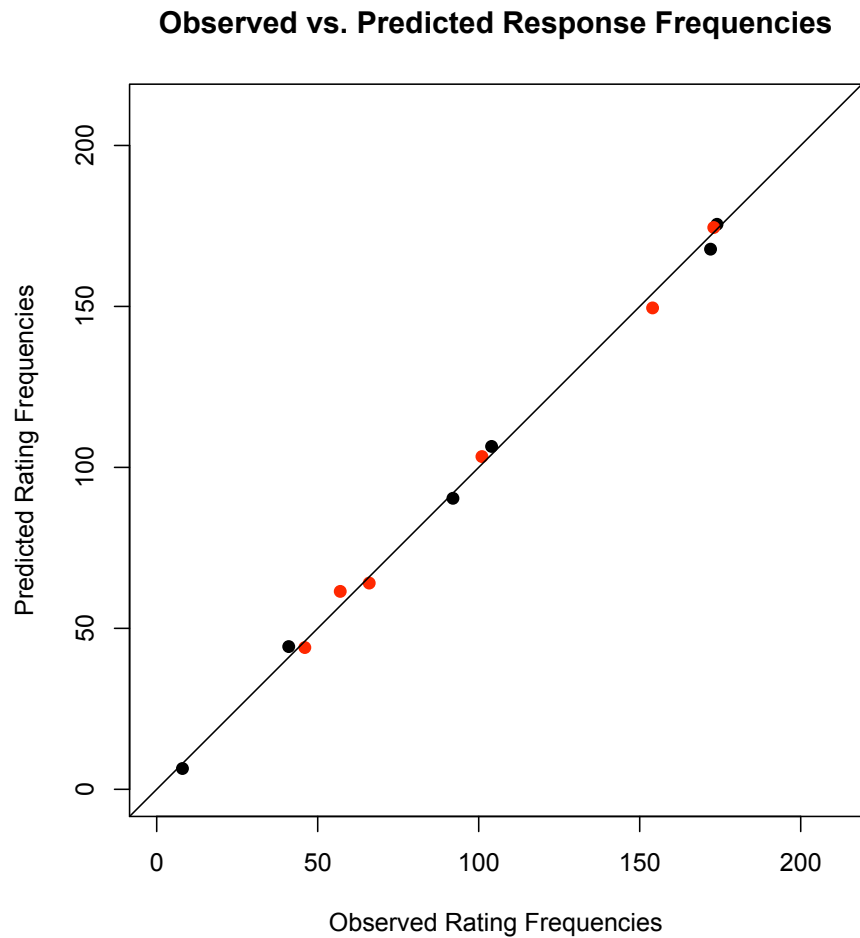
#### Signal Detection Model with Decision Criteria



**Observed vs. Predicted Response Frequencies**

X-axis: Obs\_Freq0, Obs\_Freq1

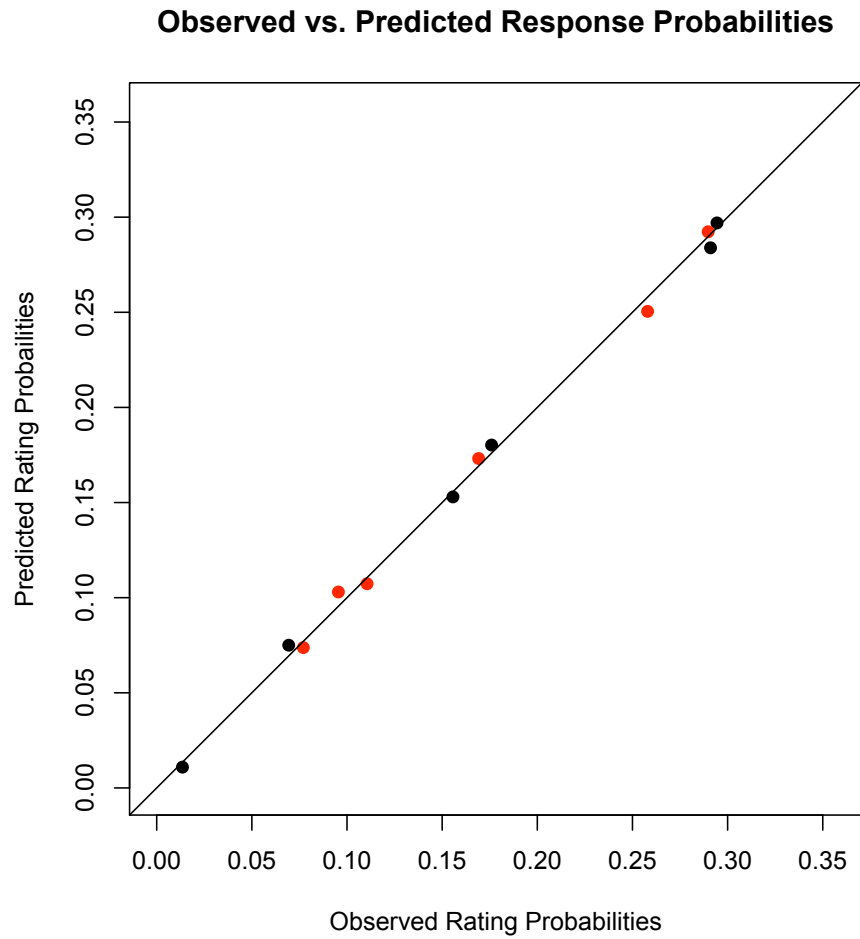
Y-axis: Pred\_Freq0, Pred\_Freq1



**Observed vs. Predicted Response Probabilities**

X-axis: Obs\_Prob0, Obs\_Prob1

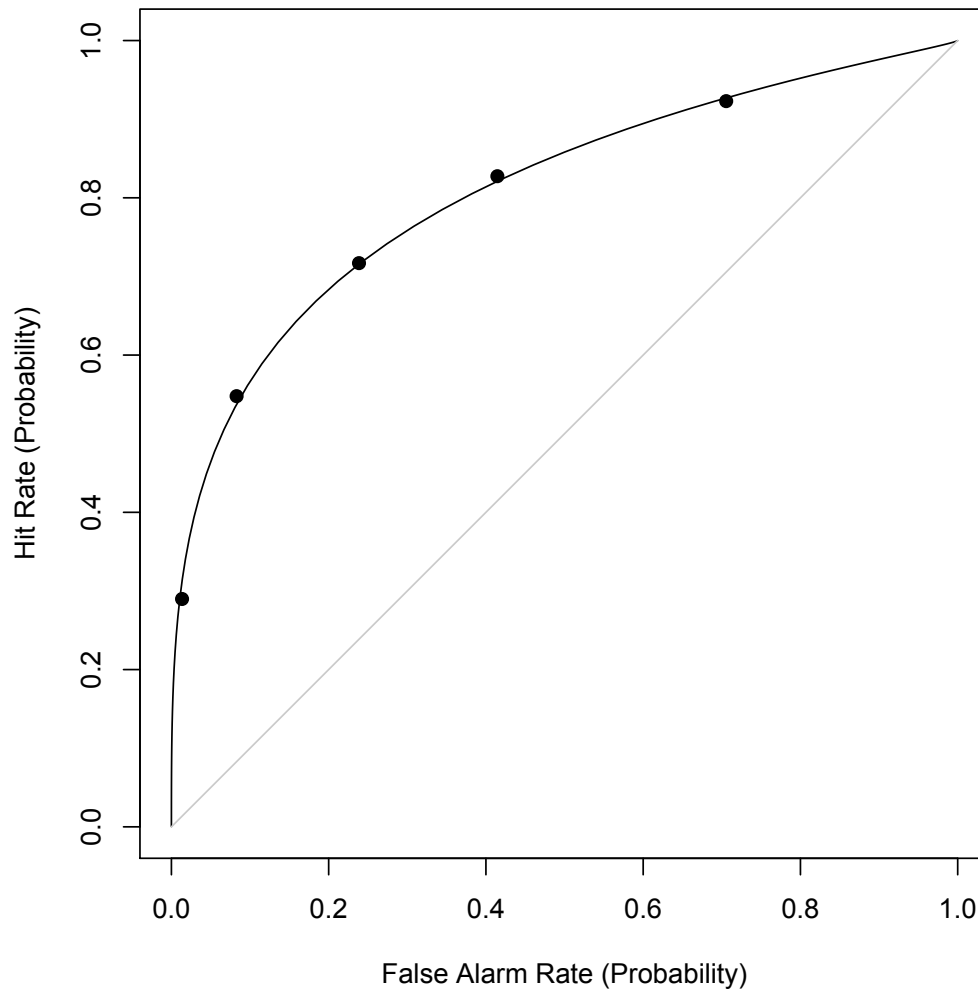
Y-axis: Pred\_Prob0, Pred\_Prob1



**Probability Receiver Operating Characteristic**

X-axis: Obs\_pYs0, Pred\_pYs0 (false alarm rates)

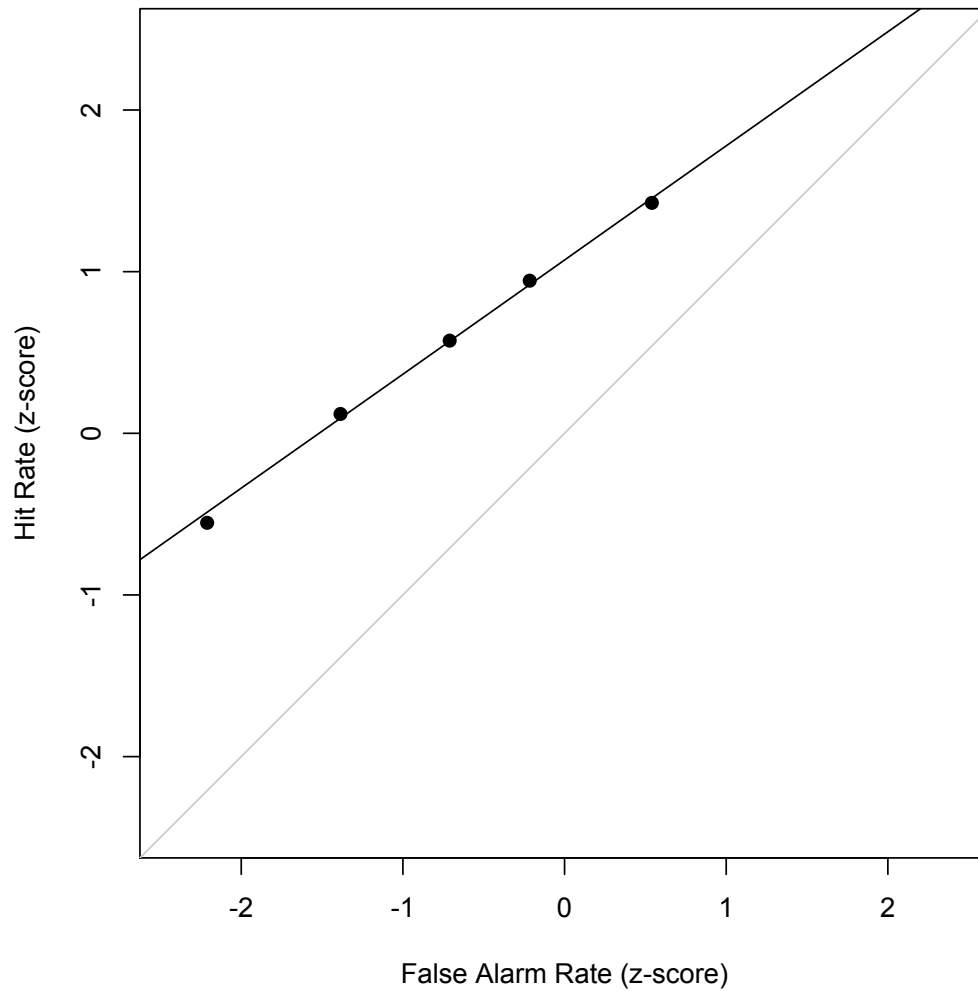
Y-axis: Obs\_pYs1, Pred\_pYs1 (hit rates)

**Observed vs. Predicted ROC**

**Z-Score Receiver Operating Characteristic**

X-axis: Obs\_zYs0, Pred\_zYs0 (z-score of false alarm rates)

Y-axis: Obs\_zYs1, Pred\_zYs1 (z-score of hit rates)

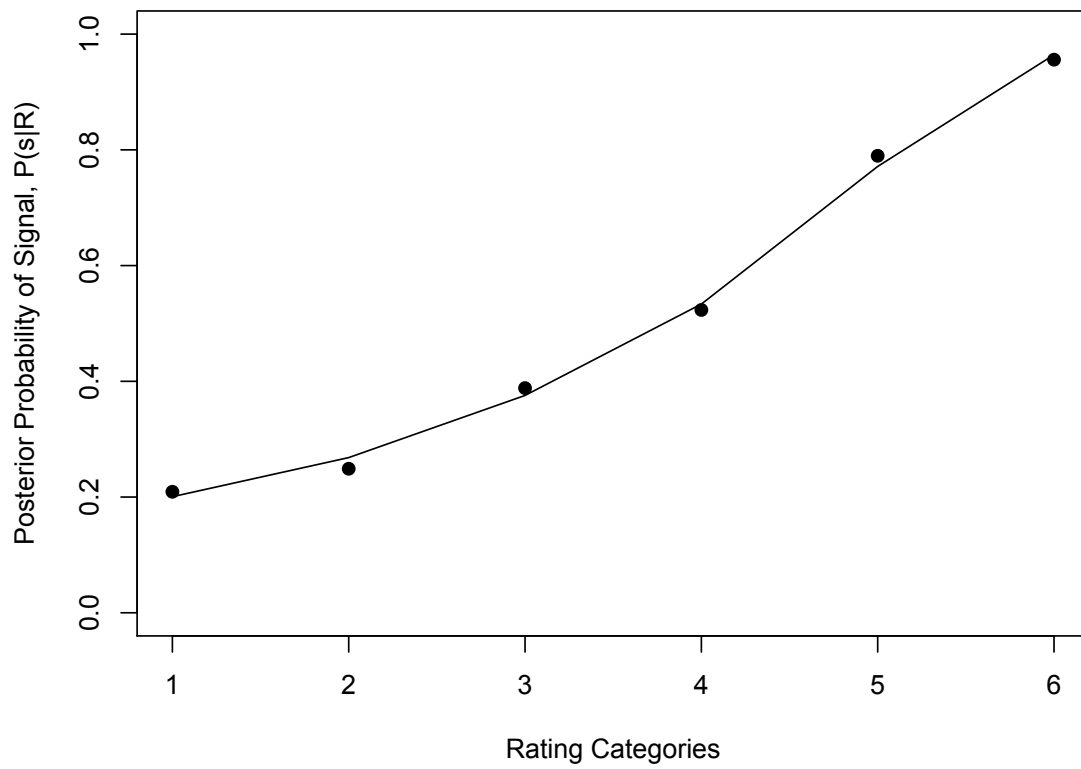
**Observed vs. Predicted z-score ROC**



**Posterior Signal Probability for each response category**

X-axis: Rating\_Category

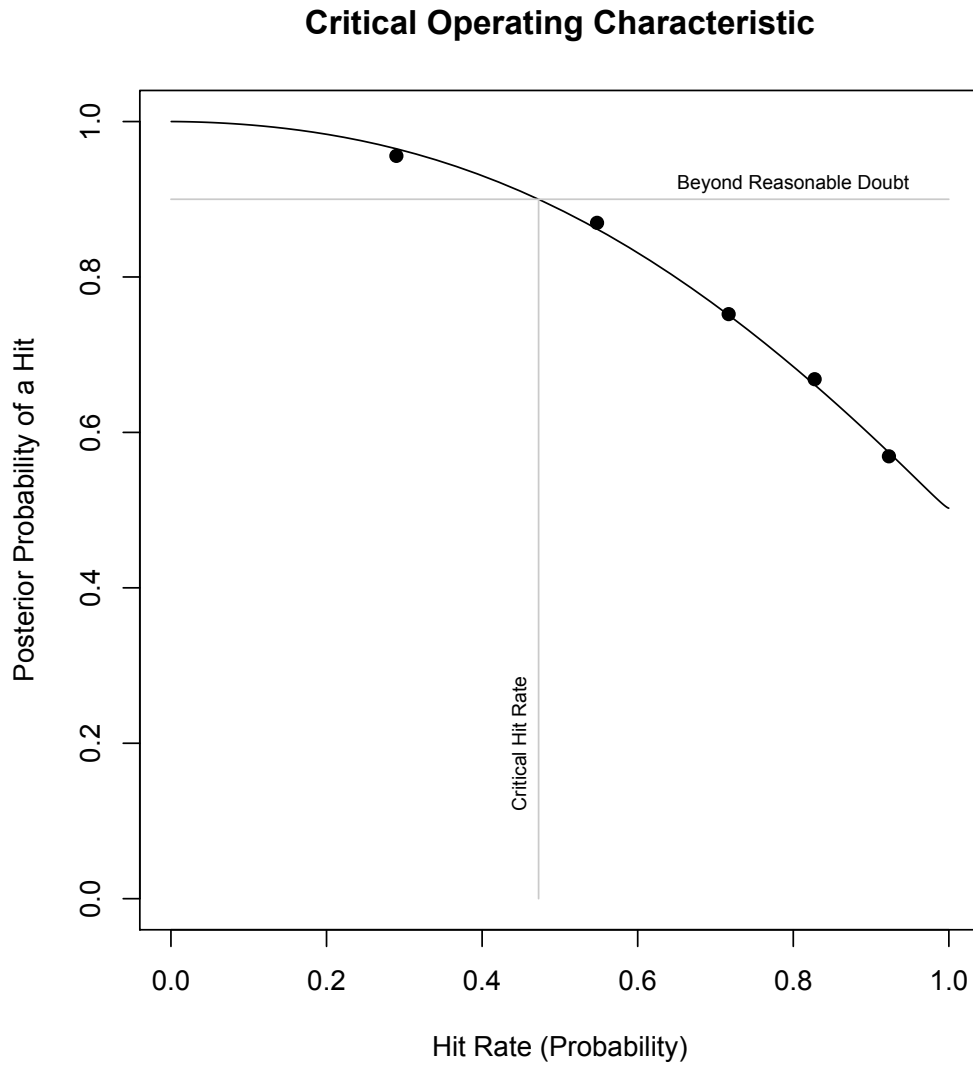
Y-axis: Obs\_pSr1, Pred\_pSr1 (posterior probabilities)

**Posterior Probability of Signal**

**Critical Operating Characteristic** (Harvey, 1992)

X-axis: Obs\_pYs1, Pred\_pYs1 (hit rates)

Y-axis: Obs\_pHitR1, Pred\_pHitR1 (posterior probability of hit)



### Further Reading

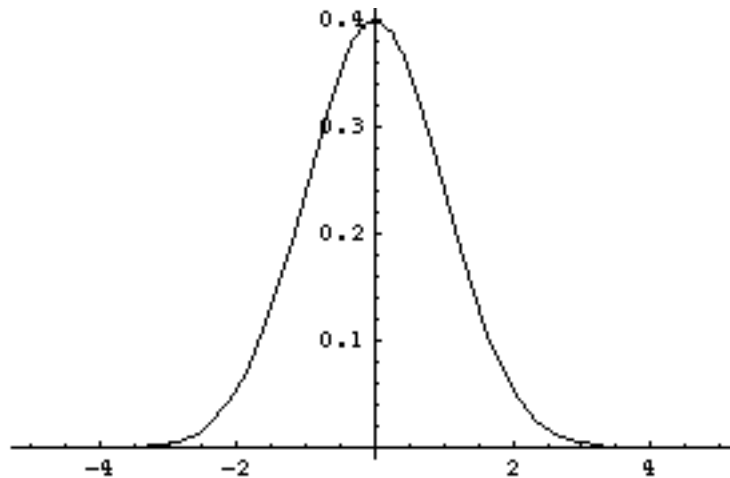
In addition to the references cited above I encourage you to consult the following references to develop a deep understanding of signal detection theory. The starting point should be the two seminal papers published in 1961 by John Swets (Swets, 1961; Swets, et al., 1961) and David Krantz's review of detection theories (Krantz, 1969). The classic book is by Green and Swets (1966/1974). Other books are also valuable (Baird & Noma, 1978; Egan, 1975; Laming, 1986; Macmillan & Creelman, 1991, 2005; McNicol, 1972; Torgerson, 1958; Wickens, 2002). Don't forget to read Louis Thurstone's original papers that lay the foundation for modern detection theory (Thurstone, 1927a, 1927b). And don't forget just how smart Fechner was (Link, 1994). A clear discussion of various indices of sensitivity within the framework of signal detection theory is given by Simpson and Fitter (1973). The use of reaction time as a surrogate for confidence is reviewed by Pike (1973). A more sophisticated model that incorporates reaction time is described by Ratcliff (2009)

The literature on the application of SDT to various applied areas is vast. Here are a few references: (Harvey, 1992; Harvey, Hammond, Lusk, & Mross, 1992; Swets, 1986a, 1986b; Swets, Dawes, & Monahan, 2000; Swets & Pickett, 1982).

Finally, here is a grab bag of papers that I have found valuable over the years: (DeCarlo, 1998, 2002; Dorfman & Alf, 1968, 1969; Dorfman, et al., 1973; Dorfman & Berbaum, 1995; Dorfman, Berbaum, & Lenth, 1995; Dorfman, Berbaum, & Metz, 1992; Dorfman, et al., 1997; Green, 1960; Green & Birdsall, 1978; Healy & Kubovy, 1978; Koppell, 1977; Macmillan & Creelman, 1990; Ogilvie & Creelman, 1968; Ratcliff & Starns, 2009; Rosner & Kochanski, 2009; Shimamura & Wickens, 2009; Simpson & Fitter, 1973; Swets, 1973, 1992, 1996, 1964; Tanner & Swets, 1953, 1954; Van Zandt, 2000; Wixted, 2007).

## Appendix I: Gaussian Probability Distribution

The Gaussian distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970a, Chapter 13):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:**  $f(x : \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-0.5 \left( \frac{x - \mu}{\sigma} \right)^2}$

**Cumulative Distribution Function:**  $F(x : \mu, \sigma) = \frac{1 + \operatorname{erf}\left(\frac{x - \mu}{\sigma \sqrt{2}}\right)}{2}$

**Quantile Function:**  $Q(p : \mu, \sigma) = \mu + \sigma \sqrt{2} \left[ \operatorname{erf}^{-1}(0, 2p - 1) \right]$

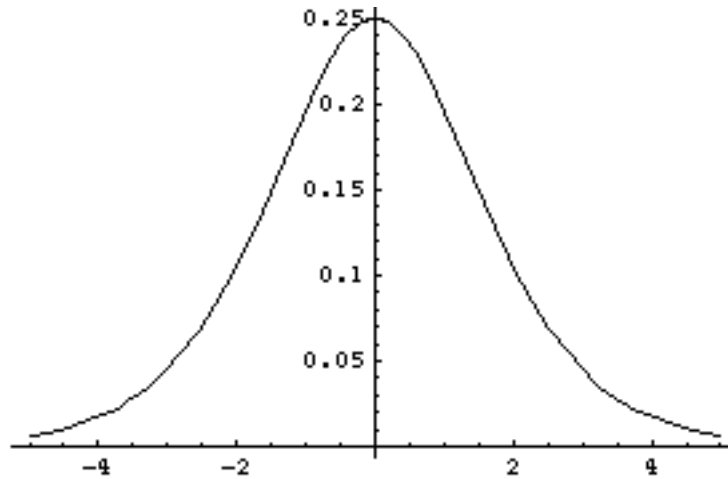
**Mean:**  $\mu$

**Standard Deviation:**  $\sigma$

**Decision Criteria:** The decision criteria are expressed in standard deviation units of the first signal distribution.

## Appendix II: Logistic Probability Distribution

The native logistic distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970b, Chapter 22):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:** 
$$f(x : \alpha, \beta) = \frac{e^{-\left(\frac{x-\alpha}{\beta}\right)}}{\beta \left(1 + e^{-\left(\frac{x-\alpha}{\beta}\right)}\right)^2}$$

**Cumulative Distribution Function:** 
$$F(x : \alpha, \beta) = \frac{1}{1 + e^{-\left(\frac{x-\alpha}{\beta}\right)}}$$

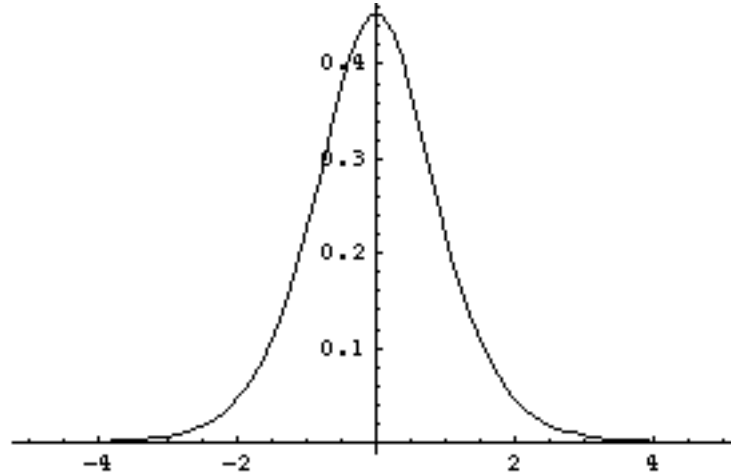
**Quantile Function:** 
$$Q(p : \alpha, \beta) = \alpha + \beta (\ln(p) - \ln(1 - p))$$

**Mean:**  $\alpha$

**Standard Deviation:** 
$$\frac{\beta \pi}{\sqrt{3}}$$

**Decision Criteria:** The decision criteria are expressed in units of  $\beta$  of the first signal distribution.

The standardized logistic distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970b, Chapter 22):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:** 
$$f(x; \mu, \sigma) = \frac{e^{-\left(\frac{\pi(x-\mu)}{\sqrt{3}\sigma}\right)} \pi}{\sqrt{3} \left(1 + e^{-\left(\frac{\pi(x-\mu)}{\sqrt{3}\sigma}\right)}\right)^2} \sigma$$

**Cumulative Distribution Function:** 
$$F(x; \mu, \sigma) = \frac{1}{1 + e^{-\left(\frac{\pi(x-\mu)}{\sqrt{3}\sigma}\right)}}$$

**Quantile Function:** 
$$Q(p; \mu, \sigma) = \mu + \frac{\sigma\sqrt{3}}{\pi} (\ln(p) - \ln(1-p))$$

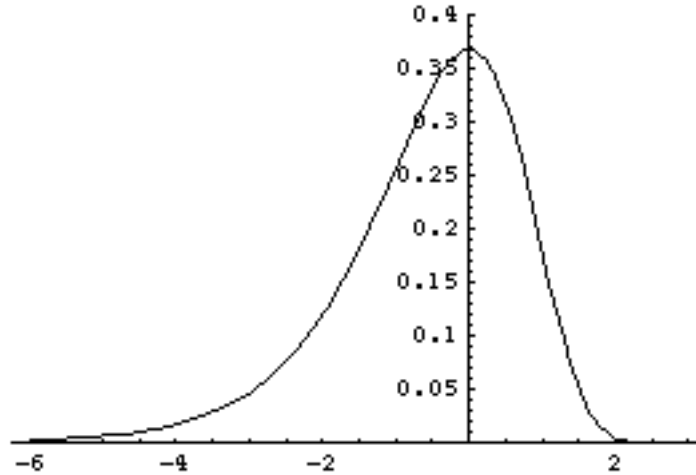
**Mean:**  $\mu$

**Standard Deviation:**  $\sigma$

**Decision Criteria:** The decision criteria are expressed in units of  $\sigma$  of the first signal distribution.

### Appendix III: Extreme Minimum Value Probability Distribution

The native extreme minimum value distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970a, Chapter 21):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:** 
$$f(x : \alpha, \beta) = \frac{e^{-e^{\left(\frac{x-\alpha}{\beta}\right)}} + \left(\frac{x-\alpha}{\beta}\right)}{\beta}$$

**Cumulative Distribution Function:** 
$$F(x : \alpha, \beta) = 1 - e^{-e^{\left(\frac{\alpha-x}{\beta}\right)}}$$

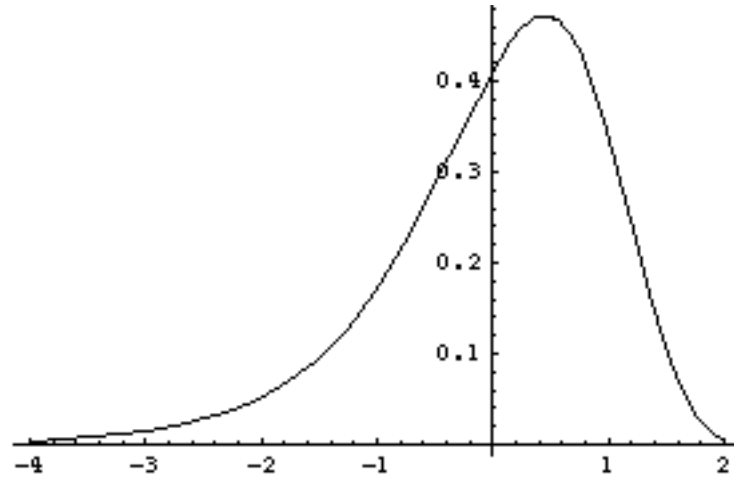
**Quantile Function:** 
$$Q(p : \alpha, \beta) = \alpha + \beta \ln\left(\ln\left(\frac{1}{1-p}\right)\right)$$

**Mean:**  $\beta \text{ EulerGamma} - \alpha$

**Standard Deviation:** 
$$\frac{\beta \pi}{\sqrt{6}}$$

**Decision Criteria:** The decision criteria are expressed in units of  $\beta$  of the first signal distribution.

The standardized extreme minimum value distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970a, Chapter 21):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:**

$$f(x; \mu, \sigma) = \frac{e^{-e^{\left( \frac{\pi \left( \mu - \frac{\sqrt{6} \text{EulerGamma } \sigma}{\pi} + x \right)}{\sqrt{6} \sigma} \right)}} + e^{\left( \frac{\pi \left( \mu - \frac{\sqrt{6} \text{EulerGamma } \sigma}{\pi} + x \right)}{\sqrt{6} \sigma} \right)}}}{\sqrt{6} \sigma \pi}$$

**Cumulative Distribution Function:**  $F(x; \mu, \sigma) = 1 - e^{-e^{\left( \frac{\pi \left( \mu - \frac{\sqrt{6} \text{EulerGamma } \sigma}{\pi} + x \right)}{\sqrt{6} \sigma} \right)}}$

**Quantile Function:**  $Q(p; \mu, \sigma) = -\mu + \frac{\sigma \sqrt{6}}{\pi} \left( \text{EulerGamma} + \ln \left( \ln \left( \frac{1}{1-p} \right) \right) \right)$

**Mean:**  $\mu$

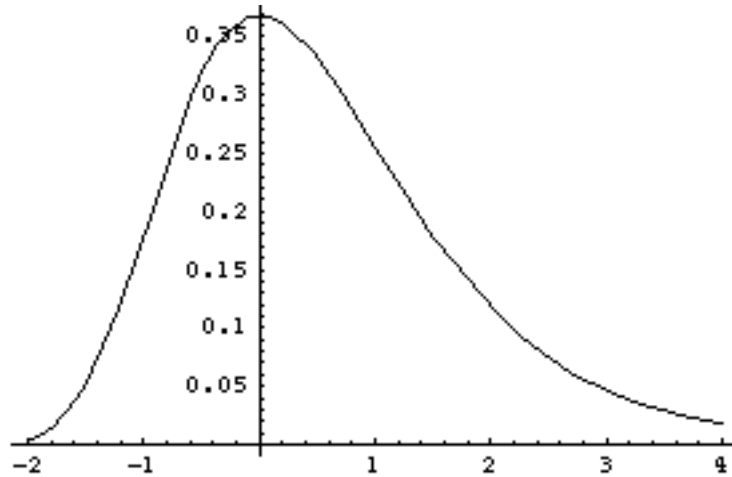
**Standard Deviation:**  $\sigma$

**Decision Criteria:** The decision criteria are expressed in units of  $\sigma$  of the first signal distribution.



### Appendix IV: Extreme Maximum Value Probability Distribution

The native extreme maximum value distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970a, Chapter 21):



**Domain:**  $-\infty$  to  $+\infty$

**Probability Density Function:** 
$$f(x : \alpha, \beta) = \frac{e^{-e^{\left(\frac{\alpha-x}{\beta}\right)} + \left(\frac{\alpha-x}{\beta}\right)}}{\beta}$$

**Cumulative Distribution Function:** 
$$F(x : \alpha, \beta) = e^{-e^{\left(\frac{x-\alpha}{\beta}\right)}}$$

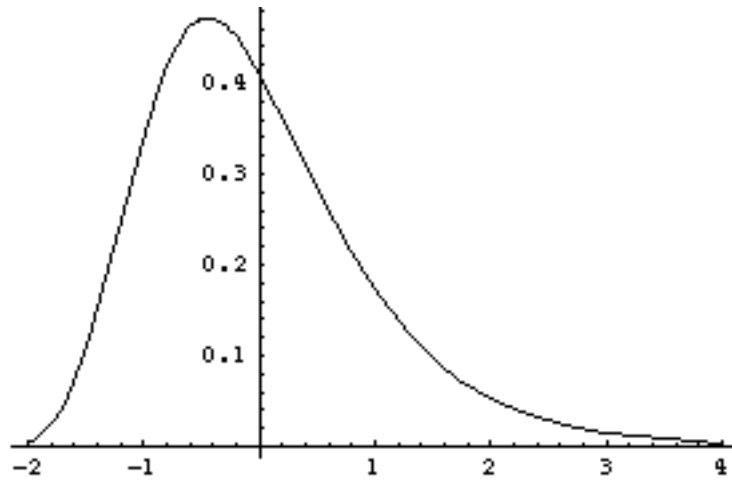
**Quantile Function:** 
$$Q(p : \alpha, \beta) = \alpha - \beta \ln\left(\ln\left(\frac{1}{1-p}\right)\right)$$

**Mean:**  $\alpha + \beta \text{ EulerGamma}$

**Standard Deviation:** 
$$\frac{\beta \pi}{\sqrt{6}}$$

**Decision Criteria:** The decision criteria are expressed in units of  $\beta$  of the first signal distribution.

The standardized extreme maximum value distribution fitted by RscorePlus has the following properties (Johnson & Kotz, 1970a, Chapter 21):



**Domain:**  $-\infty$  to  $+\infty$

$$\text{Probability Density Function: } f(x : \mu, \sigma) = \frac{e^{-e^{\frac{\pi \left( \mu - \frac{\sqrt{6} \text{ EulerGamma } \sigma}{\pi} - x \right)}{\sigma \sqrt{6}}} + \frac{\pi \left( \mu - \frac{\sqrt{6} \text{ EulerGamma } \sigma}{\pi} - x \right)}{\sigma \sqrt{6}}}}{\pi \sigma \sqrt{6}}$$

$$\text{Cumulative Distribution Function: } F(x : \mu, \sigma) = e^{-e^{\frac{\pi \left( -\mu + \frac{\sqrt{6} \text{ EulerGamma } \sigma}{\pi} + x \right)}{\sigma \sqrt{6}}}}$$

$$\text{Quantile Function: } Q(p : \mu, \sigma) = \mu - \frac{\sqrt{6} \sigma \text{ EulerGamma}}{\pi} - \frac{\sqrt{6} \sigma \ln \left( \ln \left( \frac{1}{p} \right) \right)}{\pi}$$

**Mean:**  $\mu$

**Standard Deviation:**  $\sigma$

**Decision Criteria:** The decision criteria are expressed in units of  $\sigma$  of the first signal distribution.

### Appendix V: High Threshold Model

The high threshold model (HTM) assumes that the sensory process contains a sensory threshold: a stimulus level above which the sensory process produces an output and below which the sensory process produces no output (Krantz, 1969). The PDF and CDF's below are not true probability density function and cumulative distribution functions because the area under the PDF is not 1.0 and the CDF does not range from 0.0 to 1.0 (it ranges from  $p$  to 1.0)

**Domain:** 0 to 1

**Probability Density Function:**  $1.0 - p$

**Cumulative Distribution Function:**  $F(x : p) = p + (1.0 - p) x$

**Quantile Function:**  $Q(x : p) = \frac{hr - p}{1.0 - p}$

**Mean:**

**Standard Deviation:**

**Decision Criteria:** The decision criteria are expressed in probability of guessing.

**Appendix VI: R-code for making SDT Graphs**

```

# SDT_Graphs.R
#
# This R script file creates various signal detection graphs
# using the *_grf.txt files created by RscorePlus.
# The snippets below are keyed to the examples in the
# RscorePlus User's Manual
#
# Lewis O. Harvey, Jr.
# Department of Psychology and Neuroscience
# University of Colorado
# Boulder, Colorado, USA
#
# 1 February 2009
#
#
# To create any graph from the graphical output of RscorePlus
# there are three steps to follow:
#
# 1. Read the contents of the file into an R data frame;
# 2. Select the subset of data to plot, typically a single data set;
# 3. Issue plot commands using the appropriate pairs of variables.

#####
# Read the file into a data frame and select a subset
sn <- 1 # data set number; change
as appropriate
fn <- file.choose() # choose the file to read
df.all <- read.delim(fn, header = TRUE) # reads tab delimited file
df <- subset(df.all, data_set == sn) # select a data set

#####
# Figure out how many criteria and how many signal conditions
nsig <- df$nsig[1] # number of signals
nrat <- df$nrat[1] # number of ratings
ncrt <- nrat - 1 # number of criteria

par (mfcol = c(4 ,2)) # plot all graphs in one figure

```

```

# Plot 1
#####
# Observed vs. Predicted ROC (probability coordinates)
ind0 <- match("Obs_pYs0", names(df)) # find s0 observed p(Y|s) column
ind1 <- match("Pred_pYs0", names(df)) # find s0 predicted p(Y|s) column
par(pty="s") # make square

# plot s1 ROC
plot(df[, ind0], df[, ind0 + 1], type = "p", pch = 19,
      xlim = c(0, 1),
      ylim = c(0, 1),
      xlab = "False Alarm Rate (Probability)",
      ylab = "Hit Rate (Probability)",
      main = "Observed vs. Predicted ROC")
lines(df[, ind1], df[, ind1 + 1])
lines(c(0,1), c(0,1), col = "gray", lty = "solid")

# plot the ROCs (s2...)
colors <- rainbow(nsig - 1)
if (nsig > 2){
  for (i in 2:(nsig-1)){
    points(df[, ind0], df[, ind0+i], type = "p",
           pch = 19, col = colors[i-1])
    lines(df[, ind1], df[, ind1+i], col = colors[i-1])
  }
}

par(pty="m")

```

```

# Plot 2
#####
# Observed vs. Predicted ROC (z-score probability coordinates)
ind0 <- match("Obs_zYs0", names(df)) # find s0 observed z(Yls) column
ind1 <- match("Pred_zYs0", names(df)) # find s0 predicted z(Yls) column
lim <- 1.1 * max(abs(df[, ind0:(ind0+nsig-1)]), na.rm = TRUE)

par(pty="s") # make square

# plot s1 ROC
plot(df[, ind0], df[, ind0 + 1], type = "p", pch = 19,
      xlim = c(-lim, lim),
      ylim = c(-lim, lim),
      xlab = "False Alarm Rate (z-score)",
      ylab = "Hit Rate (z-score)",
      main = "Observed vs. Predicted z-score ROC")
lines(df[, ind1], df[, ind1 + 1])
abline(0, 1, col = "gray", lty = "solid")

# plot the ROCs (s2...)
colors <- rainbow(nsig - 1)
if (nsig > 2){
  for (i in 2:(nsig-1)){
    points(df[, ind0], df[, ind0+i], type = "p",
           pch = 19, col = colors[i-1])
    lines(df[, ind1], df[, ind1+i], col = colors[i-1])
  }
}

par(pty="m")

```

```

# Plot 3
#####
# Observed vs. Predicted Response Frequencies
ind0 <- match("Obs_Freq0", names(df)) # find s0 response frequency
column
ind1 <- ind0 + (2 * nsig) - 1
max.freq <- 1.2 * max(df[, ind0:ind1], na.rm = TRUE)
par(pty="s") # make square

# plot s0 response frequencies
plot(df[, ind0], df[, ind0 + nsig], type = "p", pch = 19,
      xlim = c(0, max.freq),
      ylim = c(0, max.freq),
      xlab = "Observed Rating Frequencies",
      ylab = "Predicted Rating Frequencies",
      main = "Observed vs. Predicted Response Frequencies")

# plot the remaining response frequencies (s1...)
colors <- rainbow(nsig - 1)
for (i in (ind0+1):(ind0+nsig-1)){
  points(df[, i], df[, i+nsig], type = "p",
         pch = 19, col = colors)}
abline(0, 1, col = "black", lty = "solid")
par(pty="m")

```

```

# Plot 4
#####
# Observed vs. Predicted Response Probabilities
ind0 <- match("Obs_Prob0", names(df)) # find s0 response probability
column
ind1 <- ind0 + (2 * nsig) - 1
max.prob <- 1.2 * max(df[, ind0:ind1], na.rm = TRUE)
par(pty="s") # make square

# plot s0 response probabilities
plot(df[, ind0], df[, ind0 + nsig], data = df, type = "p", pch = 19,
      xlim = c(0, max.prob),
      ylim = c(0, max.prob),
      xlab = "Observed Rating Probabilities",
      ylab = "Predicted Rating Probabilities",
      main = "Observed vs. Predicted Response Probabilities")

# plot the remaining response probabilities (s1...)
colors <- rainbow(nsig - 1)
for (i in (ind0+1):(ind0+nsig-1)){
  points(df[, i], df[, i+nsig], data = df, type = "p",
         pch = 19, col = colors)
}
abline(0, 1, col = "black", lty = "solid")
par(pty="m")

```



```

# Plot 5
#####
# Observed and Predicted posterior probability of signal given response
ind0 <- match("Rating_Category", names(df)) # find the rating scale
ind1 <- match("Obs_pSr1", names(df))       # find posterior prob of
signal 1
ind2 <- match("Pred_pSr1", names(df))      # find posterior prob of
signal 1

par(pty="m") # don't make square

# plot s1 posterior
plot(df[, ind0], df[, ind1], type = "p", pch = 19,
     ylim = c(0, 1),
     xlab = "Rating Categories",
     ylab = "Posterior Probability of Signal, P(s|R)",
     main = "Posterior Probability of Signal")
lines(df[, ind0], df[, ind2])

# plot the remaining posterior probabilities (s2...)
colors <- rainbow(nsig - 1)
if (nsig > 2){
  for (i in 1:(nsig-1)){
    points(df[, ind0], df[, ind1+i], type = "p",
           pch = 19, col = colors[i])
    lines(df[, ind0], df[, ind1+i], col = colors[i])
  }
}

par(pty="m")

```

```

# Plot 6
#####
# Signal Detection Model with Decision Criteria

par(pty="m")

# find the maximum density value for setting y-axis
ind0 <- match("Pred_Density0", names(df)) # find s0 density column
ind1 <- ind0 + nsig - 1
max.density <- 1.2 * max(df[, ind0:ind1], na.rm = TRUE)

# plot s0 probability density function
plot(df[, ind0] ~ Xc, data = df, type = "l",
      xlim = c(min(df$Xc, na.rm = TRUE), max(df$Xc, na.rm = TRUE)),
      ylim = c(-0.03, max.density),
      xlab = "Decision Axis (standard deviations of S0)",
      ylab = "Probability Density",
      main = "Signal Detection Model with Decision Criteria")

# plot the remaining probability density functions (s1...)
colors <- rainbow(nsig-1)
for (i in 1:(nsig-1)){
  lines(df[, ind0+i] ~ Xc, data = df, type = "l", col = "black")}
abline(h = 0, col = "gray", lty = "solid")

# extract the criterion coordinates
crit <- subset(df, Criteria != is.na(Criteria))[c("Xc", "Xc_se",
"Criteria")]
xc <- df$Xc[1:ncrt]
# plot the decision criterion markers
for (i in 1:ncrt){
  segments(crit$Xc[(2*i)-1], crit$Criteria[(2*i)-1], crit$Xc[2*i],
crit$Criteria[2*i], col="red")}
# plot the standard error of each criterion marker
for (i in 1:ncrt){
  segments(crit$Xc[2*i]-crit$Xc_se[2*i], crit$Criteria[2*i],
crit$Xc[2*i]+crit$Xc_se[2*i], crit$Criteria[2*i], col = "red")}

# label the response categories
xc.lab <- rep(0, nrat)
xc.lab[2:ncrt] <- xc[1:ncrt-1] + diff(xc)/2
xc.lab[1] <- xc[1] - (diff(xc)/2)[1]
xc.lab[nrat] <- xc[ncrt] + (diff(xc)/2)[ncrt-1]
text(xc.lab, -0.025, labels = 1:nrat, cex = 0.7)

par(pty="m")

```

```

# Plot 7
#####
# Observed and Predicted critical operating characteristic
ind0 <- match("Obs_pYs1", names(df))      # find the observed hit rate
ind1 <- match("Obs_pHitXc1", names(df))    # find the observed
posterior of hit
ind2 <- match("Pred_pYs1", names(df))      # find the predicted hit rate
ind3 <- match("Pred_pHitXc1", names(df))    # find the predicted
posterior of hit

par(pty="s") # make square

# plot s1 posterior
plot(df[, ind0], df[, ind1], type = "p", pch = 19,
      xlim = c(0, 1),
      ylim = c(0, 1),
      xlab = "Hit Rate (Probability)",
      ylab = "Posterior Probability of a Hit",
      main = "Critical Operating Characteristic")
lines(df[, ind2], df[, ind3])
lines(c(0, 1), c(0.9, 0.9), col = "gray")
text(0.8, 0.92, "Beyond Reasonable Doubt", cex = 0.7)

inter <- approx(df[, ind2], df[, ind3], n=1000)
ind4 <- match(TRUE, inter$y < 0.9)
if (!is.na(ind4)) {
  xc <- inter$x[ind4]
  yc <- inter$y[ind4]
  lines(c(xc, xc), c(0, yc), col = "gray")
  text(xc-.025, .2, "Critical Hit Rate", cex = 0.7, srt = 90)
}

# plot the remaining posterior probabilities (s2...)
colors <- rainbow(nsig - 1)
if (nsig > 2){
  for (i in 1:(nsig-1)){
    points(df[, ind0+i], df[, ind1+i], type = "p",
           pch = 19, col = colors[i])
    lines(df[, ind2+i], df[, ind3+i], col = colors[i])
    inter <- approx(df[, ind2+i], df[, ind3+i], n=1000)
    ind4 <- match(TRUE, inter$y < 0.9)
    if (!is.na(ind4)) {
      xc <- inter$x[ind4]
      yc <- inter$y[ind4]
      lines(c(xc, xc), c(0, yc), col = colors[i])
    }
  }
}
par(pty="m")

```

### A Signal Detection Theory Reference List

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6), 716–723.
- Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.
- DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, 3(2), 186–205.
- DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, 109(4), 710–721.
- Dorfman, D. D., & Alf, E., Jr. (1968). Maximum Likelihood Estimation of Parameters of Signal Detection Theory: a Direct Solution. *Psychometrika*, 33(1), 117–132.
- Dorfman, D. D., & Alf, E., Jr. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating method data. *Journal of Mathematical Psychology*, 6(3), 487–496.
- Dorfman, D. D., Beavers, L. L., & Saslow, C. (1973). Estimation of signal detection theory parameters from rating-method data: A comparison of the method of scoring and direct search. *Bulletin of the Psychonomic Society*, 1(3), 207–208.
- Dorfman, D. D., & Berbaum, K. S. (1995). Degeneracy and discrete receiver operating characteristic rating data. *Academic Radiology*, 2(10), 907–915.
- Dorfman, D. D., Berbaum, K. S., & Lenth, R. V. (1995). Multireader, multicase receiver operating characteristic methodology: A bootstrap analysis. *Academic Radiology*, 2(7), 626–633.
- Dorfman, D. D., Berbaum, K. S., & Metz, C. E. (1992). Receiver Operating Characteristic Rating Analysis: Generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*, 28(9), 723–731.
- Dorfman, D. D., Berbaum, K. S., Metz, C. E., Lenth, R. V., Hanley, J. A., & Dagga, H. A. (1997). Proper receiver operating characteristic analysis: the bigamma model. *Academic Radiology*, 4(2), 138–149.
- Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. New York: Academic Press.
- Green, D. M. (1960). Psychoacoustics and signal detection theory. *Journal of the Acoustical Society of America*, 32, 1189–1203.
- Green, D. M., & Birdsall, T. G. (1978). Detection and Recognition. *Psychological Review*, 85, 192–206.
- Green, D. M., & Swets, J. A. (1966/1974). *Signal detection theory and psychophysics* (A reprint, with corrections of the original 1966 ed.). Huntington, NY: Robert E. Krieger Publishing Co.
- Harvey, L. O., Jr. (1992). The critical operating characteristic and the evaluation of expert judgment. *Organizational Behavior & Human Decision Processes*, 53(2), 229–251.
- Harvey, L. O., Jr., Hammond, K. R., Lusk, C. M., & Mross, E. F. (1992). Application of signal detection theory to weather forecasting behavior. *Monthly Weather Review*, 120(5), 863–883.

- Hautus, M. J., & Lee, A. J. (1998). The dispersions of estimates of sensitivity obtained from four psychophysical procedures: implications for experimental design. *Perception & Psychophysics*, 60(4), 638–649.
- Hays, W. L. (1973). *Statistics for the Social Sciences* (2nd ed.). New York: Holt, Rinehart and Winston.
- Healy, A. F., & Kubovy, M. (1978). The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition*, 6(5), 544–553.
- Johnson, N. L., & Kotz, S. (1970a). *Continuous univariate distributions-1*. New York: John Wiley & Sons.
- Johnson, N. L., & Kotz, S. (1970b). *Continuous univariate distributions-2*. New York: John Wiley & Sons.
- Koppell, S. (1977). Decision latencies in recognition memory: A signal detection theory analysis. *Journal of Experimental Psychology: Human Learning & Memory*, 3(4), 445–457.
- Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, 76(3), 308–324.
- Laming, D. (1986). *Sensory analysis*. New York: Academic Press.
- Link, S. W. (1994). Rediscovering the past: Gustav Fechner and signal detection theory. *Psychological Science*, 5(6), 335–340.
- Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and “nonparametric” indexes. *Psychological Bulletin*, 107(3), 401–413.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185–199.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11, 431–441.
- McNicol, D. (1972). *A primer of signal detection theory*. London: George Allen & Unwin.
- Ogilvie, J. C., & Creelman, C. D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, 5, 377–391.
- Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, 80(1), 53–68.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing* (2nd ed.). Cambridge, England: Cambridge University Press.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116(1), 59–83.

- Rosner, B. S., & Kochanski, G. (2009). The Law of Categorical Judgment (Corrected) and the interpretation of changes in psychophysical performance. *Psychological Review*, 116(1), 116-128.
- Shimamura, A. P., & Wickens, T. D. (2009). Superadditive memory strength for item and source recognition: The role of hierarchical relational binding in the medial temporal lobe. *Psychological Review*, 116(1), 1-19.
- Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, 80(6), 481-488.
- Swets, J. A. (1961). Is there a sensory threshold? *Science*, 134(3473), 168-177.
- Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, 182(4116), 990-1000.
- Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99(2), 181-198.
- Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99(1), 100-117.
- Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47(4), 522-532.
- Swets, J. A. (1996). *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Swets, J. A. (Ed.). (1964). *Signal Detection and Recognition by Human Observers*. New York: John Wiley & Sons, Inc.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1-26.
- Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: methods from signal detection theory*. New York: Academic Press.
- Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68(5), 301-340.
- Tanner, W. P., Jr., & Swets, J. A. (1953). *A new theory of visual detection* (Technical Report No. 18): Electronic Defense Group, University of Michigan.
- Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6), 401-409.
- Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology*, 38, 368-389.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26(3), 582-600.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.
- Wixted, J. T. (2007). Dual-Process Theory and Signal-Detection Theory of Recognition Memory. *Psychological Review*, 114(1), 152-176.