SDT Assistant V.2.0 - 2021

Table of Contents

SDT Assistant V.2.0 - 2021

ROC Analysis

Algorithms and Other Details

General Suitability

The ROC analysis component of SDT Assistant is capable of fitting ROC curves to data collected using the rating-response protocol in combination with the following psychophysical tasks: (1) Yes-No (also called A-Not A); (2) two-alternative forced-choice (2AFC; also called paired comparison); (3) 2AFC with Reminder (2AFCR; also called fixed-reference duo-trio); (4) Same-Different (note, standard detection-theoretic version with four stimulus sequences); (5) Yes-No with Reminder (also called A-Not A with Reminder). SDT Assistant is not suitable for the ROC analysis of multi-point data collected in multiple experiments using a binary-response protocol.

Models Fitted

The following is a brief description of the models available for each task, the method used to find the best-fitting parameters given the data, and brief details of the results reported by SDT Assistant. Note that names for decision strategies are different within psychology and sensory evaluation. In general, the β strategy is interchangeable with the likelihood-ratio strategy, the τ strategy is interchangeable with the difference strategy, and the COD (comparison of distances) strategy is interchangeable with the absolute-difference strategy.

Unequal-Variance Normal Model for Yes-No or A Not-A Ratings (β Strategy)

Brief Description: This is the standard detection-theoretic model that is the most commonly described model in the literature. See Green and Swets (1966) for a description of this model.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure. Intercept, slope, and criteria are fitted.

Results Reported: Apart from the parameters of the model, d_a (Simpson & Fitter, 1973) is also provided.

Equal-Variance Normal Model for Yes-No or A Not-A Ratings (β Strategy)

Brief Description: This is a restricted version of the unequal-variance normal model in which the standard deviations of the two normal distributions are equal. See Green and Swets (1966) for a description of this model.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure. Intercept and criteria are fitted.

Results Reported: Apart from the parameters of the model, d' (equal to the Intercept), is provided separately.

Equal-Variance Normal Model for 2AFC Ratings (β/τ Strategy)

Brief Description: This is the standard version of the 2AFC task. See Hautus, Macmillan, and Creelman (2022) for a description of the model.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure. Intercept and criteria are fitted.

Results Reported: Apart from the parameters of the model, d' (equal to the Intercept), is provided separately.

Equal-Variance Normal Model for Yes-No with Reminder Ratings (β & τ Strategies)

Brief Description: This is the standard version of the Yes-No Reminder task. Two versions of the model are available; one for the β strategy, and the other for the τ strategy. See Hautus, van Hout, and Lee (2009) for a description of these models.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure. Intercept and criteria are fitted.

Results Reported: Apart from the parameters of the model, d' (equal to the Intercept), is provided separately.

Equal-Variance Normal Model for 2AFC with Reminder Ratings (β, τ & COD Strategies)

Brief Description: This is the standard version of the 2AFC with Reminder task. Two versions of the model are available; one for the β and τ strategies – which have identical ROC curves and are henceforth combined and referred to as the β/τ strategy – and the other for the comparison of distances (COD) strategy. See Hautus, van Hout, and Lee (2009) for a description of these models.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure are provided for the β/τ strategy. For the COD strategy, a slower minimization procedure is employed. Intercept and criteria are fitted.

Results Reported: Apart from the parameters of the model, d', is provided separately.

Equal-Variance Normal Model for Same-Different Ratings (β & τ Strategies)

Brief Description: This is the standard version of the same-different task. Two versions of the model are available; one for the β strategy, and the other for the τ strategy. See Hautus et al. (2022) for a description of these models.

Fitting Method: Maximum Likelihood estimates of parameters provided using a quadratic convergence procedure are provided for both strategies. Parameters estimated are the distance between the bivariate-normal distributions (d') and the criteria.

Results Reported: The parameters of the model, including d', are reported.

Maximum Likelihood Estimation and Maximization Algorithms

All models are fitted to data using maximum likelihood estimation. The general Log Likelihood equation employed is that given by Dorfman and Alf (1969; Equation 3) for the rating-response protocol. Many of the models are fitted using a quadratically-convergent method; in this case, the

Newton-Raphson method. The model for the 2AFCR task with the COD strategy is currently fitted using a simplex method.

Fitting an ROC Curve

The use of the software is (sort of and becoming more) intuitive.

1. Enter ratings data in the grid. Some data are present when you run the software, to make it easier to get a feel for how it works. These are the data from the example given below.
2. Select the model from the drop-down box.
3. Click on Process, and then on Process Current Grid.

Sometimes it is desirable to process data from a text file. An example file is included with the software. The general format looks as follows:

```
        6
Jack
       10    20    30    40    50    60   210
       60    50    40    30    20    10   210
Jill
        7    12    45    67    89   120   340
      120    89    67    45    12     7   340
```

The first row contains the number of ratings. The file can then repeat as many blocks of three rows as desired. Row 1 is a label, which can be blank. Rows 2 & 3 contain the ratings. Rows 2 and 3 are terminated with the row total. Note that it is easy to export data from Excel and other programs in this format. Once a text file is constructed, adjust the settings in the software, and then click Process, then Process File. The rest is self-explanatory. An output file will be constructed that has the same name as your data file, but with '.prn' appended. This is a standard text file. Note that only one analysis result is ever presented on the screen. New results overwrite old results on the screen. The output file contains all of the individual analyses.

Results and their Interpretation: An Example

To illustrate ROC analysis in SDT Assistant, the data from Green and Swets (1966 , p. 102) is analysed. These data correspond to the analysis for Observer 1 in Dorfman and Alf (1969) and are also analysed as Example 1 in DeCarlo (2003). The data originate from Swets, Tanner, and Birdsall (1961) but were not published in that paper in raw form.

The following annotated output will provide assistance with interpreting results:

```
        Unequal-Variance Yes-No (LR/Beta) Rating Model
```

*** This is the model that was fitted to the data.*

```
Original Data Matrix...

        46         57         66        101        154        173     Total: 597
       174        172        104         92         41          8     Total: 591
```

** These are the original data.

```
Original Proportions Matrix...

         –       .923       .827       .717       .548       .290
         –       .706       .415       .239       .083       .014
```

** These are the original false-alarm (bottom) and hit (top) rates.

```
Processed Data Matrix...

       46.0       57.0       66.0      101.0      154.0      173.0
      174.0      172.0      104.0       92.0       41.0        8.0
```

** This is the adjusted data matrix to which the model was fitted. Redundant rating categories (those with a frequency of zero in both positions) are eliminated, and 0.5 is added to the rightmost cells if the log-linear correction is applied. Note that no adjustments have been made in this case.

```
Processed Proportions Matrix...

         –       .923       .827       .717       .548       .290
         –       .706       .415       .239       .083       .014
```

** These are the false alarm and hit rates that stem from the adjusted matrix.

```
Fitted Parameters...

-5.331E-01 2.042E-01 7.099E-01 1.366E+00 2.294E+00 1.072E+00 7.058E-01
```

** These are the best-fitting parameters. Estimates of the criteria are presented first (one less than the number of rating categories processed; so five in this case), followed by the intercept, and the slope (relevant for this model only and not reported for others).

```
Derivatives...

3.950E-14  2.952E-13  2.173E-13  8.589E-12 -1.248E-11  4.060E-12 -1.429E-11
```

** The derivatives of the log likelihood function for each of the parameters. These should be very small if a maximum has actually been found.

```
Worst Derivative Magnitude: 1.429E-11
```

*** This is the largest absolute value from the previous list. It provides a one-look diagnostic for whether the fitting process terminated acceptably. A small value, say below 1E-3, should be ok. A large value indicates a degenerate fit.*

```
Convergence: 4.510E-06
```

*** This is the sum over parameters of the absolute value of the difference between each final parameter value and that from the preceding iteration. The value should be less than the convergence criterion if the fitting process was successful.*

```
Variance-Covariance Matrix...

2.870E-03  1.659E-03  1.220E-03  7.612E-04  4.013E-05  1.450E-03  6.425E-04
           2.452E-03  1.911E-03  1.539E-03  1.245E-03  1.397E-03  2.080E-04
                      2.771E-03  2.450E-03  2.579E-03  1.332E-03 -2.563E-04
                                 4.527E-03  5.462E-03  1.101E-03 -1.243E-03
                                            1.265E-02  2.687E-04 -3.477E-03
                                                       4.214E-03  1.362E-03
                                                                  1.947E-03
```

*** The asymptotic variance-covariance matrix for the parameters. Variances are on the diagonal. Off-diagonal values are covariances. The ordering is the same as in the parameters list above.*

```
Initial Values

Intercept=       1.075E+00
Slope=           7.188E-01
```

*** Starting values used in the minimization process.*

```
Final Values

Intercept=       1.072E+00     Var= 4.214E-03     SE= 6.492E-02
Slope=           7.058E-01     Var= 1.947E-03     SE= 4.413E-02

da=              1.239E+00     Var= 4.452E-03     SE= 6.672E-02
```

*** Final estimates for the major parameters, usually only d' (technically the Intercept), but Intercept and Slope for the model fitted here. Also, for this model, the sensitivity measure d_a is provided. Variance and standard errors as extracted and calculated (e.g., for d_a via the delta method) from the variance-covariance matrix above.*

```
Criteria...

xc =            -5.331E-01   2.042E-01   7.099E-01   1.366E+00   2.294E+00
```

```
VAR =              2.870E-03   2.452E-03   2.771E-03   4.527E-03   1.265E-02
SE =               5.358E-02   4.952E-02   5.264E-02   6.728E-02   1.125E-01
```

** *Final estimates for the criterial parameters, specified as the location of the criteria on the x-axis of the decision space, and sometimes referred to as x_c. In some literature and software packages, these are referred to as the z(k). Note that for the Equal-Variance Normal model, criteria specified relative to the cross-over point of the distributions of evidence, commonly called c, are also provided.*

```
Area measures...

RIndex (Ag or p(A)) =  8.031E-01   Var=   1.633E-04   SE=   1.278E-02
AUC =                  8.095E-01   Var=   1.645E-04   SE=   1.283E-02
```

** *Area measures of sensitivity are reported. The RIndex is referred to as p(A) or A_g in the psychological literature. The error measures for this index are those given by Hanley and McNeil (1982). They do not assume continuous distributions. The Area Under Curve, which is the entire area under the ROC curve is often referred to as A_z in the literature.*

```
Fit Statistics...

G2=                1.48
X2=                1.50

DF=                   3

p(G2)=         6.865E-01
p(X2)=         6.829E-01

LL=             -1920.98
AIC=             3855.97
BIC=             3891.53
```

** *Various statistics associated with goodness-of-fit: The likelihood ratio version of chi-square (G2), and the traditional chi-square (X2). DF is the number of degrees of freedom, and the two probabilities are based on the chi-square distribution. These can be interpreted as the probability that the data arose from the model given that the model is correct. The maximized log likelihood (LL), the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are provided for comparison of un-nested models.*

```
Notifications...

Max Iterations =        50
```

** *The maximum number of iterations permitted to reach the convergence criterion.*

```
Iterations =             5
```

*** Number of iterations used to reach convergence. If this number is the same as Max Iterations then it is unlikely that the fitting process was successful. It is also likely that the value give for Convergence will be relatively large.*

```
Convergence Crit. =          1.000E-07
```

*** This is the criterion level below which the algorithm seeks to reduce residuals. The value is currently fixed and not user adjustable.*

```
Bootstrapped Estimators Based on 10000 Resamples:

Bootstrapped Variance-Covariance Matrix...
2.894E-03  1.615E-03  1.290E-03  8.500E-04  3.964E-04  1.403E-03  5.029E-04
           2.456E-03  1.942E-03  1.569E-03  1.441E-03  1.422E-03  1.691E-04
                      2.853E-03  2.543E-03  2.878E-03  1.452E-03 -2.856E-04
                                 4.766E-03  6.066E-03  1.256E-03 -1.347E-03
                                            1.426E-02  4.405E-04 -3.750E-03
                                                       4.098E-03  1.291E-03
                                                                  1.989E-03
```

*** The bootstrapped variance-covariance matrix for the parameters. Variances are on the diagonal. Off-diagonal values are covariances. The ordering is the same as in the parameters list above. Bootstrapping is only available for the first four models in the drop-down box.*

```
Fitted Parameters
-5.331E-01 2.042E-01 7.099E-01 1.366E+00 2.294E+00 1.072E+00  7.058E-01
Bootstrapped Parameters
-5.347E-01 2.048E-01 7.127E-01 1.370E+00 2.305E+00 1.073E+00  7.057E-01
Bias
-1.627E-03 6.357E-04 2.770E-03 3.872E-03 1.073E-02 5.878E-04 -1.031E-04
```

*** Fitted parameters as per top section and bootstrapped parameters (mean of bootstrap distribution). The difference between these is the bias. The ordering is the same as in the parameters list above. Note that there is minimal bias in this case.*

```
RIndex from Data      8.031E-01
RIndex Bootstrapped   8.029E-01        Var= 2.198E-04      SE= 1.482E-02

Bias                 -2.684E-04
```

*** RIndices obtained from the original data and from bootstrapping (mean of bootstrap distribution). Bootstrapping provides an estimate of the variance and standard errors of the RIndex. Note that there is minimal bias in this case.*

```
AUC from Data          8.095E-01
AUC Bootstrapped       8.093E-01           Var= 2.307E-04          SE= 1.519E-02
Bias                  -1.676E-04
```

** *Area Under Curve obtained from the original data and from bootstrapping (mean of bootstrap distribution). Bootstrapping provides an estimate of the variance and standard errors of the AUC. Note that there is minimal bias in this case.*

Miscellaneous other Results

For the Equal-Variance Normal model SDT Assistant also reports a "non-parametric" version of d' which is derived from the RIndex (or A_g or $p(A)$). The following would appear below the Area Measures section and in the bootstrapping section.

```
Approximate d' from RIndex...
d'app =                1.206E+00

d'app from Data        1.206E+00
d'app Bootstrapped     1.206E+00           Var= 4.192E-03          SE= 6.475E-02
Bias                   1.391E-04
```

Other Options

Print Output
Clicking on Process and then Print Output results in the current output being sent to the default printer. If you wish to send the output to a different printer, you need to select that printer as the default printer in Windows <u>before</u> starting SDT Assistant. Printer selection will be built into SDT Assistant in a future version.

Bootstrap Replications
Clicking on Settings and then Bootstrap Replications will bring up a submenu that allows you to select the number of replications to perform. Options include None (default), 100, 300, 1000, 3000, and 10000. Bootstrapping is only available for the first four models in the drop-down box. ROC fitting of other models will automatically default this setting back to None.

Log-Linear Correction
Clicking on Settings and then on Log-Linear Correction instructs SDT Assistant to add a value of 0.5 to the cell frequencies for the highest rating and to add a value of 1 to the row totals. During calculation of false-alarm and hit rates, the value of 0.5 will be propagated to all frequencies, essentially applying a log-linear correction to each proportion. This is a frequently-used correction for binary response data, but has not been thoroughly investigated for rating data. Notwithstanding this, the correction is very useful when the number of trials is relatively small. It frequently allows an estimate of ROC parameters in cases where the standard fitting process fails. If this option is selected, the effect on the data to which the model is fitted can be viewed in the output, in the 'Processed Data Matrix' and 'Processed Proportions Matrix' sections.

Output Style

The default output style, illustrated in the output example above, is Standard Output. The other options are Single-Line Output and DFA Output.

Single-Line Output is useful for extracting certain key values from an analysis and presenting it on a line. This allows for large data files to be processed with the analysis for each data set to be on a line. Very useful for later analysis in another program. The output on each line takes the form

```
    2.032E-13    3   1.072E+00    7.099E-01    -1921.00    1.51    1.51
```

where the values are the Worst Derivative Magnitude, DF, Intercept, Slope, Log Likelihood, Chi-Square and G2 respectively.

DFA Output is designed to make analysis of the Double-Faced Applicability test (Kim, Hopkinson, van Hout, & Lee, 2017a, 2017b) easier to perform. For this output to be valid, bootstrapping must be enabled and the equal-variance model fitted. Each line of output will take the form

```
    4.096E-01       2.120E-02       1.456E-01       1.242E-01       6.950E-01
```

where the values correspond to d'app, and its variance, standard error, and lower and upper 95% confidence intervals respectively.

ROC Graphs

SDT Assistant produces a plot of each ROC curve that is fitted. This is primarily for the purpose of eyeballing what the data look like on the best-fitting curve. The graph can be printed, but will be of relatively low quality. Additional features will be added in updates to the software. The default setting is for this feature to be on. It can be turned off/on by clicking on Settings and then on View ROCs. The following figure gives an idea of what is produced by this function.
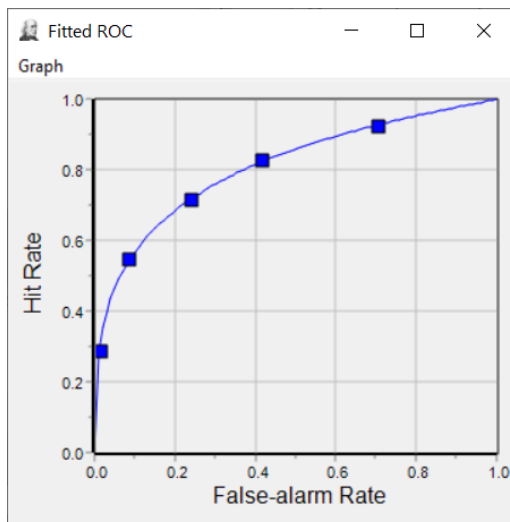
Curve Generator

The curve generator in SDT Assistant allows you to generate and save an [x, y] dataset that specifies a curve. The dataset can then be imported into software that can perform graphing and used to specify a curve for plotting. The curve generator panel is shown in the following picture and the function of each panel is described below.

Normal Density Parameters

This panel allows the user to specify the *Mean* and *Standard Deviation* of a normal density function. The *Maximum X* and *Minimum X* values specify the domain over which function values are generated. This domain is split into *Values* equal-spaced data points. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved.

ROC (UEV Normal) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the Yes-No task; sometime called the A-Not A task. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate. It can produce both unequal- and equal-variance normal ROC curves. For equal-variance normal ROC curves, set *Slope* to 1. For this case *Intercept* is equal to *d'*. A value of *Slope* different from 1 produces an unequal-variance normal ROC curve. In this case *d'* is not a relevant quantity, and the *Intercept* is literally the intercept of the ROC curve when plotted on inverse-normal coordinates, or coordinates of z-scores. On these coordinates the ROC curve becomes linear, with a slope of *Slope* and an intercept of *Intercept*. This is also true of

equal-variance normal ROC curves, in which case the slope on inverse-normal coordinates is 1 and the intercept is equal to d'.

ROC (Same-Different) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the standard same-different task. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate. The only parameter that can be specified is d'. The panel can produce ROC curves for both the difference (τ) and likelihood ratio (β) models of the same different task. The model is selectable by the radio buttons. There is also an option to *Mirror* the difference ROC curve about the negative diagonal of the unit square. This feature allows production of curves that define a hit in terms of same stimuli or in terms of different stimuli. Both versions can be found in the literature.

ROC (ABX) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the standard ABX task. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. This task is equivalent to the balanced-reference duo-trio task. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate. The only parameter that can be specified is d'. The panel can produce ROC curves for both the difference (τ) and likelihood ratio (β) models of the ABX task. The model is selectable by the radio buttons. It is worth noting that in this case the τ strategy is equivalent to the comparison of distances strategy.

ROC (2AFC) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the two-alternative forced-choice (2AFC) task. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate. The only parameter that can be specified is d'. For the 2AFC task the difference (τ) and likelihood ratio (β) models yield the same outcomes and hence a single curve represents both models.

ROC (2AFCR) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the two-alternative forced-choice with reminder (2AFCR) task, with either the absolute difference (Absolute Difference – Comparison of Distances) strategy or difference-likelihood ratio strategy. The 2AFCR task is structurally identical to the fixed-reference duo-trio task. In the food science literature, the COD strategy is usually assumed for the duo-trio task and hence this option generates the ROC curve for the traditional fixed-reference duo-trio task. ROC curves for the difference (τ) or likelihood ratio (β) decision strategies for 2AFCR are equivalent. These ROC curves are identical to that for the 2AFC task, with the same value of d'. The only parameter that can be specified is d'. Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate.

ROC (YNR) Parameters

This panel generates a [false-alarm, hit rate] dataset that defines an ROC curve suitable for data collected using the Yes-No with Reminder task; sometimes called the A-Not A with Reminder task.

Selecting *Generate* will open a dialog asking for a file name to which the dataset will be saved. The dataset will consist of *Values* pairs equally spaced in terms of the false-alarm rate. The only parameter that can be specified is d'. The panel can produce ROC curves for both the difference (τ) and likelihood ratio (β) models of the Yes-No with Reminder task. The model is selectable by the radio buttons. ROC curves for the likelihood ratio strategy in Yes-No with Reminder are equivalent to those from the Equal-Variance model for the Yes-No task.

d' Calculator

The d' Calculator, available in two versions on the *Binary* menu, is specifically for the estimation of d' from tasks that employ the binary-response protocol.

d' from F and H

The first version of the calculator, *d' from F and H*, can estimate d' for all of the models listed for the ROC Analysis component. In addition, it can estimate d' for the ABX task (sometimes called the variable-reference duo trio task).

More useful than the estimates of d' are the exact estimates of the standard error (SE) of d'. These estimates are not based on asymptotic theory, which is the basis of most published tables. Instead they are based of the calculation of every possible value of d' that can be attained with a given number of trials, and the probability that each of those values of d' will arise. Full details can be found in Miller (1996) or Hautus and Lee (1998).

When the false-alarm or hit rates are zero or one it is not feasible to calculate d'. In most cases, the best estimate is that d' is infinite. To avoid such intractability in practice, and to recognize that the estimation of very large values of d' is statistically unsound, researchers employ a correction that is applied to the false-alarm and hit rates prior to estimating d'. Details of these corrections can be found in Miller (1996), Hautus (1995), Hautus et al. (2022) and elsewhere.

SDT Assistant offers three options for the correction to apply: (1) do nothing; (2) the log-linear correction (always add 0.5 to the obtained frequencies and 1.0 to the total on which they are based); and (3) the 1/(2N) correction (only if the false-alarm and/or hit rate is equal to zero or one, change it to 1/(2N) or 1-1/(2N) respectively, where N is the total number of trials upon which the false-alarm or hit rate is based). The first option does not allow the calculation of the SE using exact methods because there is always a non-zero probability of obtaining a false-alarm or hit rate of zero or one, and hence obtaining $d' = \pm\infty$. Of the remaining two options, the log-linear correction leads to estimates of d' that predictably underestimate the true value of d'. In most situations, the log-linear correction also leads to the least statistical bias. There are 'better' corrections than these (Hautus & Lee, 2006) that may be implemented in a later version of SDT Assistant.

Estimating d' from Data

An example of the use of this module is shown in the following figure.

This case illustrates a Yes-No (A-Not A) task with 60 trials in total and 30 for each stimulus type. The judge obtained 21 hits and 8 false alarms. No correction has been applied so the corrected rates are the same as the original ones. In this case the false-alarm and hit rates are those traditionally used; 8/30 and 21/30. d' is calculated in the usual manner as $z(H)-z(F)$, which in this case is equal to 1.147. Because no correction has been used, standard errors cannot be calculated using the exact approaches adopted here. One could adopt an asymptotic approximation (e.g., Gourevitch and Galanter (1967) for the Yes-No task), but for less than 100 trials, this approach is error prone (Miller, 1996). Additionally, if the judge had obtained 0 or 30 false alarms or hits – which one can expect to happen from time to time even when true d' is a reasonably small number like 1 – d' could not be calculated and nor could its standard error.

The next figure shows the same data, but this time using the log-linear correction.

This approach adds 0.5 to the obtained number of false alarms and hits, and 1.0 to the number of trials using each stimulus, before determining d'. In this case $F = 8.5/31 = 0.2742$ and $H = 21.5/31 = 0.6935$. Subtracting z-scores gives $d' = 1.106$. This means that d' can always be calculated, even for the case where the judge obtains 0 or 30 false alarms or hits. It also allows the calculation of the exact standard error of d'.

To do this, the assumption is made that the estimated value of *F* and *H* are the best estimates of the true values, as estimated proportions are unbiased estimators of population proportions. If this is the case, when an experiment is run with 60 trials as done in the example, the expected value of *d*′ is E(*d*′) = 1.098. In other words, if you ran a large number of experiments of this nature and determined the average of all of the estimates of *d*′ obtained, that average would be 1.098. It would, in this case, be slightly different from the true value. Thus, there is statistical bias in the estimates of *d*′.

VAR(*d*′) is the variance of all of the estimates of *d*′ that you may collect. It is therefore a reflection of the statistical uncertainty in the data that you have collected. SE(*d*′) is the standard error of *d*′ estimates; therefore the square root of VAR(*d*′). It is a standard error, and not a standard deviation, because it shows the variability around the *mean* of the attainable values of *d*′.

While VAR(*d*′) and SE(*d*′) indicate variability around E(*d*′), the mean-square error of *d*′, MSE(*d*′) indicates variability around the true value of *d*′. If there is no statistical bias, and E(*d*′) = true *d*′, then MSE(*d*′) = VAR(*d*′). MSE(*d*′) cannot be less than VAR(*d*′). As statistical bias gets larger, so does MSE(*d*′). The difference between VAR(*d*′) and MSE(*d*′) can serve as a warning against trusting statistical inferences drawn using the standard error of *d*′. If this difference is relatively large then your test is leaning towards assessing differences not from the true value of *d*′, but from an incorrect biased estimate of *d*′. For a more detailed account of this topic, including asymptotic estimates of standard error, see Hautus et al. (2022, pp. 303-310).

Finally, some confidence intervals are presented. These are simple confidence intervals estimated on the basis of the normal distribution of standard errors. This assumption is known to be incorrect. A more exact solution will be provided in a future version of SDT Assistant.

Different Tasks

Discussion so far has been limited to the Yes-No task. SDT Assistant can determine these statistics for a range of other tasks, as described in the introductory section above. However, some of these functions run relatively slowly, even on a fast computer. This is because a large number of values of *d*′ are required to determine each standard error (and other measures), and for some tasks, the calculation of *d*′ is fairly complicated, and therefore slow. In particular, the 2AFCR (absolute difference / COD) and ABX models can take a substantial amount of time, even with a relatively small number of trials.

To work around this problem pre-calculated tables of *d*′ are used for these tasks. SDT Assistant loads the appropriate table from a file. Four factors define a file: the task (2AFCR (absolute difference / COD), ABX (difference), or ABX (likelihood ratio)), the correction applied (log-linear or 1/(2*N*)), and the number of trials on which *F* and *H* are based, which can be unequal). SDT Assistant comes supplied with tables for equal *N* (up to 100), for each task and correction. The appropriate table will be loaded when an analysis is run. If the appropriate table is not available, you will be instructed to create one using Table Generator.exe, which can be found in the same folder as SDT Assistant.

When Table Generator is run, the following panel appears.
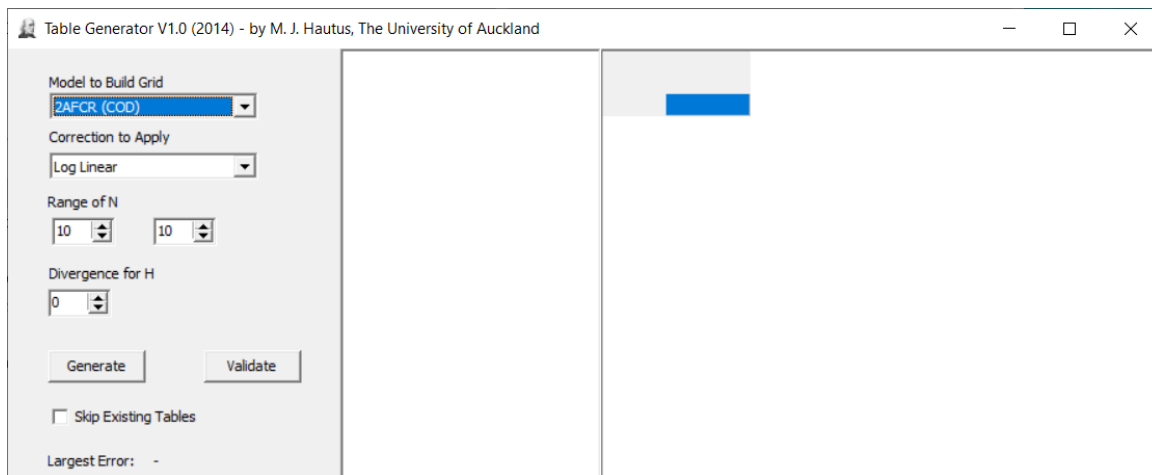
It is clear where the model and correction is specified. Other options need further explanation. 'Range of N' specifies the smallest and largest number of trials on which *F* is based. If the two values are the same, a single *d'* table will be created. If the values differ, an individual table will be created for each value of *N* in the range. 'Divergence of H' specifies the number of trials on which *H* is based, relative to the number of trials on which *F* is based. A divergence of 0 corresponds to the case of equal *N*. A divergence of 1 corresponds to the case where the number of trials upon which *H* is based is one more than that on which *F* is based. For cases of unequal *N*, two *d'* tables are generated. One for the specified case and one for the case in which the two values of *N* are interchanged. This provides an additional table, that may or may not be required, with no additional computation. Note that for each table of *d'* values (file names end in .ddat) there is an associated table of criterion values (file names end in .cdat). The criterion files are not currently used by SDT Assistant, but will be put to use in a future version.

To create tables, set the required parameters and click the Generate button. Table generation can take a long time, more so as values of *N* get larger, or as more tables are set to generate in a single session (by setting a wider range). Progress will be noted by entries in the left-hand memo area, and by the grid on the right being filled with values of *d'*. It is suggested that you practice the entire process described here using the default settings on Table Generator. The small table (100 values of *d'*) is generated fairly quickly and makes a good practice exercise.

Generated files will be found, when the program completes, in a subfolder of the GTables folder. The 'G' stands for generated. The subfolder will correspond to the task associated with the table being generated, as shown in the picture below. Never delete these folders, but feel free to delete their contents.
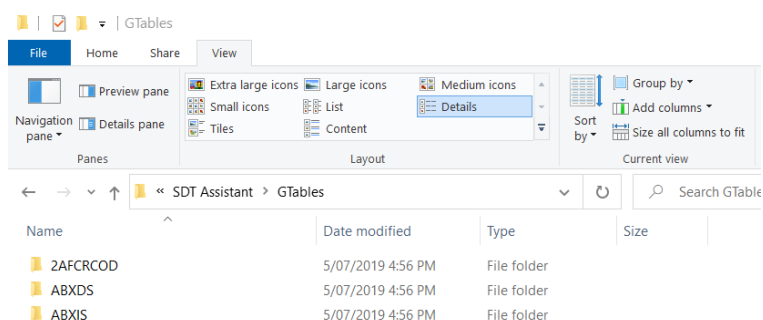
It is always a good idea to verify the files you create. This can be done by clicking the Validate button on the Table Generator panel and navigating to the appropriate file in the GTables folder. A validation process will follow, and at the end the 'largest error' in the file will be reported at the bottom left of the panel. Multiple files can be selected and processed. In this case, the largest error is across all files. It would be good if any errors greater than 1E-7 are reported to the author, together with the parameters entered into Table Generator that were used to generate the table.

Once you have generated your tables you can move (or copy) them into the production table folders that are used by SDT Assistant. These folders are inside the Tables (not GTables) folder. Move the relevant .ddat and .cdat files into the correct folder, which will have the same name as that in the GTables folder. Run SDT Assistant and attempt your analysis. Table generation has been implemented to allow for future enhancements to SDT Assistant that will allow resampling applications in real time and allow for the calculation of exact confidence intervals.

Other Features
There are two further options in this module, accessible via the Process menu item.

Process File
This option allows a text file of data to be processed without manual intervention. There are two basic file formats. The first is for calculations based on observed values.

```
S1a
0    0    1    32    32    28    4    9500
S1b
0    0    2    32    32    28    4    9500
S2a
0    0    1    32    32    19    13    9500
S2b
0    0    2    32    32    19    13    9500
```

Each data set consists of two rows, the first of which is a header label that may be left blank. The second row contains eight pieces of information. From left to right they are:

1. Basis for calculation; 0 = Observed values; 1 = True values
2. Task Type; from 0 – 8; A-Not A; A-Not AR (DS); 2AFC (LRS/DS); 2AFCR (LRS/DS); 2AFCR (COD); SD (DS); SD (LRS); ABX (DS); ABX (LRS)
3. Correction Type; 0 = None; 1 = Log Linear; 2 = 1/(2N)
4. Number of Trials that Hits are based on
5. Number of Trials that False Alarms are based on
6. Number of Hits
7. Number of False Alarms
8. Confidence Interval Percentile x 100

When the data file is selected, after clicking on Process and then Process File, each pair of lines will be processed, and the results sent to a file with the same name as the input file, but appended with '.prn'.

The format of the output is determined by the *File Output Format* setting, which can be set to either *Standard Output* or to *Single-Line Output*.

For these data with the *Standard Output* setting, the output (truncated) is…

> S1a
> Basis for calculations: Observed Values
> Task: A-Not A
> Correction: Log Linear
> Trials for H: 32
> Trials for F: 32
> Obtained H: 28
> Obtained F: 4
> H Rate (original): 8.750E-01
> H Rate (corrected): 8.636E-01
> F Rate (original): 1.250E-01
> F Rate (corrected): 1.364E-01
> d': 2.194E+00
> d' Expected: 2.175E+00
> d' Bias: -1.812E-02
> d' Variance: 1.583E-01
> d' Bias (squared): 3.282E-04
> d' MSE: 1.587E-01
> d' SE: 3.979E-01
> CI Size (x100): 9500
> CI (2-tail) Lower: 1.414E+00
> CI (2-tail) Upper: 2.974E+00
> CI (1-tail) Lower: 1.539E+00
> CI (1-tail) Upper: 2.848E+00
>
>
> S1b
> Basis for calculations: Observed Values
> Task: A-Not A
> Correction: 1/(2N)
> Trials for H: 32
> etc…

The second file format is for calculations based on hypothesized True values of *F* and *H*. The statistics produced on this setting correspond to those presented in published tables. Indeed, a data file of the type being described here can be constructed to generate statistics to be inserted in customized tables. The format is very similar to that described above, again with eight pieces of information. However, in this case item 6 is the True Hit rate, and item 7 is the True False-alarm rate. The output file is very similar to that just described.

For these same data with the Single-Line Output setting, the output (truncated) is…

| 4 | 32 | 28 | 32 | 2.194E+00 | 2.175E+00 | -1.812E-02 | 1.583E-01 | 3.979E-01 | 3.282E-04 | 1.587E-01 | 9500 | 1.414E+00 | 2.974E+00 |
| 4 | 32 | 28 | 32 | 2.301E+00 | 2.408E+00 | 1.073E-01 | 1.994E-01 | 4.465E-01 | 1.152E-02 | 2.109E-01 | 9500 | 1.426E+00 | 3.176E+00 |

The origin of each value can be determined from the complete output.

Print Output

Clicking on Process and then Print Output results in the current output being sent to the default printer. If you wish to send the output to a different printer, you need to select that printer as the default printer in Windows <u>before</u> starting SDT Assistant. Printer selection will be built into SDT Assistant in a future version.

d' from $p(c)_{max}$

The second version of the calculator, *d' from p(c)$_{max}$*, can estimate *d'* for a selection of models for which this approach is relatively common (but not necessarily recommended). The underlying consideration here is the assumption that the data were produced by an unbiased observer. Currently SDT Assistant can produce results for five tasks (see following figure). Others will be added upon request.

There are only two values to enter: the number of trials in total and the number of correct trials. All other values are calculated when the Process button is pressed.

Data may be processed from a file of the following format:

> George
> 100 60
> Sarah
> 100 80
> Peter
> 100 46

As elsewhere, each data set takes up two rows. The first is a label and the second contains the total number of trials followed by the number of correct trials. The output file looks as follows:

George	100	60	6.000E-01	3.583E-01	1.115E+00	1.976E+00	1.669E+00	1.363E+00
Sarah	100	80	8.000E-01	1.190E+00	2.355E+00	3.129E+00	2.606E+00	2.095E+00
Peter	100	46	4.600E-01	-1.420E-01	-6.766E-01	1.250E+00	1.064E+00	8.747E-01

The first three values on a row are the data provided to SDT Assistant. Then follows $p(c)$ and the five values of d' in the order shown on the panel.

Preference Models

This module conducts an analysis of preference data collected in a paired-preference experiment in which placebo pairs are employed, along with rating responses. The theory behind the analysis is provided, together with simulated data and experimental data in Zhang, Halim, Wichchukit, O'Mahony, and Hautus (2016). The form, illustrated in the following picture, is pre-populated with some of the experimental data reported by the authors.

The algorithms employed to fit the models are the same as those used for ROC fitting in SDT Assistant. You are encouraged to work through Zhang *et al* (2016) to discover the nuances of this approach.

Tables from Detection Theory: A User's Guide

Appendix 5 of Hautus et al. (2022) contains tables that are useful for conducting calculations related to the content of the book. SDT Assistant can provide most of the values within these table, and others not provided. This reduces the need to interpolate between table entries. In some instances, the values given by SDT Assistant have higher accuracy. For Table A5.3, SDT Assistant uses direct calculations that avoid the approximation that uses $z(H) - z(F)$ as the basis of calculation.

Currently, SDT Assistant does not provide values for the m-interval Oddity task. This is because the values (except for the special case of m = 3) would need to be generated by simulation, rather than calculation.

The following figure shows the form used to calculate values.

Entering values into any of the unshaded cells leads to calculation of the other values in that section of the form. For example, in the upper section one can enter either a value of probability, z-score, or likelihood. For whichever value is entered, SDT Assistant will calculate the other two values. In the special case that the likelihood is entered, both positive and negative z-scores are provided, as well as the two probabilities associated with these z-scores.

References

DeCarlo, L. T. (2003). Using the PLUM procedure of SPSS to fit unequal variance and generalized signal detection models. *Behavior Research Methods Instruments & Computers, 35*(1), 49-56. doi:10.3758/bf03195496

Dorfman, D. D., & Alf, E. (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. doi:10.1016/0022-2496(69)90019-4

Gourevitch, V., & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika, 32*(1), 25-30. doi:10.1007/BF02289402

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29-36. doi:10.1148/radiology.143.1.7063747

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods Instruments & Computers, 27*(1), 46-51. doi:10.3758/BF03203619

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. doi:10.3758/BF03206051

Hautus, M. J., & Lee, A. J. (2006). Estimating sensitivity and bias in a yes/no task. *British Journal of Mathematical & Statistical Psychology, 59*(2), 257-273. doi:10.1348/000711005x65753

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2022). *Detection Theory: A User's Guide* (3rd ed.). New York: Routledge.

Hautus, M. J., van Hout, D., & Lee, H. S. (2009). Variants of A Not-A and 2AFC tests: Signal Detection Theory models. *Food Quality and Preference, 20*(3), 222-229. doi:10.1016/j.foodqual.2008.10.002

Kim, I.-A., Hopkinson, A., van Hout, D., & Lee, H.-S. (2017a). A novel two-step rating-based 'double-faced applicability' test Part 1: Its performance in sample discrimination in comparison to simple one-step applicability rating. *Food Quality and Preference, 56*, 189-200. doi:10.1016/j.foodqual.2016.10.010

Kim, I.-A., Hopkinson, A., van Hout, D., & Lee, H.-S. (2017b). A novel two-step rating-based 'double-faced applicability' test. Part 2: Introducing a novel measure of affect magnitude (d'A) for profiling consumers' product usage experience based on Signal Detection Theory. *Food Quality and Preference, 59*, 141-149. doi:10.1016/j.foodqual.2017.02.011

Miller, J. (1996). The sampling distribution of d'. *Perception & Psychophysics, 58*(1), 65-72. doi:10.3758/bf03205476

Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin, 80*(6), 481-488. doi:10.1037/h0035203

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision-processes in perception. *Psychological Review, 68*(5), 301-340. doi:10.1037/0033-295x.68.5.301

Zhang, X. T., Halim, J., Wichchukit, S., O'Mahony, M., & Hautus, M. J. (2016). Paired Preference Tests: A signal detection based analysis with separate d ' values for segmentation. *Journal of Sensory Studies, 31*(6), 481-491. doi:10.1111/joss.12238