Behav Res (2011) 43:1023–1032
DOI 10.3758/s13428-011-0119-3

CDF-XL: computing cumulative distribution functions of reaction time data in Excel

George Houghton · James A. Grange

Published online: 30 June 2011

Abstract In experimental psychology, central tendencies of reaction time (RT) distributions are used to compare different experimental conditions. This emphasis on the central tendency ignores additional information that may be derived from the RT distribution itself. One method for analysing RT distributions is to construct cumulative distribution frequency plots (CDFs; Ratcliff, *Psychological Bulletin* 86:446–461, 1979). However, this method is difficult to implement in widely available software, severely restricting its use. In this report, we present an Excel-based program, CDF-XL, for constructing and analysing CDFs, with the aim of making such techniques more readily accessible to researchers, including students (CDF-XL can be downloaded free of charge from the Psychonomic Society's online archive). CDF-XL functions as an Excel workbook and starts from the raw experimental data, organised into three columns (Subject, Condition, and RT) on an Input Data worksheet (a point-and-click utility is provided for achieving this format from a broader data set). No further preprocessing or sorting of the data is required. With one click of a button, CDF-XL will generate two forms of cumulative analysis: (1) "standard" CDFs, based on percentiles of participant RT distributions (by condition), and (2) a related analysis employing the participant means of rank-ordered RT bins. Both analyses involve partitioning the data in similar ways, but the first uses a "median"-type measure at the participant level, while the latter uses the mean. The results are presented in three formats: (i) by participants, suitable for entry into further statistical analysis; (ii) grand means by condition; and (iii) completed CDF plots in Excel charts.

Keywords Cumulative distribution function · CDF · Bin analysis · RT distributions · Excel

Reaction time (RT) is a popular measure of cognitive ability. Measures of the central tendency of participants' RT distributions—typically the arithmetic mean, but also occasionally the median—are used to infer discrete psychological processes that have occurred in different experimental conditions. Although this method has proven to be very successful in cognitive psychology, relying solely on the central-tendency measures ignores the distribution that the central tendency has been calculated from. Some distributional information is provided in traditional analysis—spread of the distribution is often provided in the form of ranges of data and standard deviations, for example—but typically much distributional information is not considered. Analysing differences between experimental conditions throughout the whole RT distribution—rather than just analysing differences between central-tendency estimates—provides the researcher with more data to base their conclusions upon. As such, distributional analysis becomes a powerful constraint on theories and models of human cognition.

One such method for analysing RT distributions is to construct cumulative distribution frequency plots (CDFs), a method advocated in a seminal paper by Ratcliff (1979). This method, however, is computationally expensive and not easy to implement in widely available software. In this article, we present an Excel-based program for constructing and analysing CDFs, with the aim of making such

Electronic supplementary material The online version of this article (doi:10.3758/s13428-011-0119-3) contains supplementary material, which is available to authorized users.

G. Houghton (✉)
School of Psychology, University of Wales,
Bangor,
Gwynedd, Wales, UK
e-mail: g.houghton@bangor.ac.uk

J. A. Grange
School of Psychology, University of Keele,
Keele,
Staffordshire, England, UK

techniques more readily accessible to cognitive researchers. We first provide a description of the CDF method, before providing some examples of its use. Then, the program itself—CDF-XL—is described in detail. CDF-XL can be downloaded free of charge from the Psychonomic Society's online archive.

Description of the CDF method

Construction of CDFs begins by rank ordering all RTs—from fastest to slowest—for each participant and for each experimental condition. For instance, a study with $n = 40$ participants compared in three conditions will produce 120 RT rankings. Each ranking is then divided into a number of consecutive, equally sized bins (or partitions), and some measure of the central tendency of each bin is calculated. Thus, instead of a single score per condition per participant, one computes k scores, where k is the number of partitions (usually $k \leq 10$). The most frequently used form of partition measure is the percentile, for instance dividing the ranking in 10 ascending deciles. Equivalent deciles (for the same condition) are then averaged across participants, producing the CDF curve for that condition. Two or more conditions can then be compared over the whole of the RT range, rather than just at the one point (the mean).

In Fig. 1, an example CDF plot is shown displaying two experimental conditions, with the data partitioned into 10 percentiles ranging from the 5th to the 95th. This figure demonstrates the advantage of CDF analysis over standard mean-RT analysis: In this example, the mean difference between Conditions 1 and 2 is 55 ms; however, at the fastest quantile (5%), the difference is only 5 ms, and the difference increases in a linear fashion towards the slower end of the distribution, to a maximum value of 123 ms at the final (95%) quantile. Any explanation of the mean difference between the two conditions should also account for why the difference is practically absent at the faster end of the distribution, and increases linearly towards slower responses (see Grange & Houghton, 2011). In the following section, we briefly discuss some examples of the use of CDFs in psychological research and summarise the types of situation in which they appear to be most useful.

What is the use of CDF analysis? Some examples In the task-switching literature, where participants are required to rapidly switch between simple cognitive tasks, CDFs have been used to constrain theoretical models of performance. Task switching incurs an RT cost as compared to repeating tasks (called the "switch cost"; see Kiesel et al., 2010; Vandierendonck, Liefooghe, & Verbruggen, 2010, for reviews). One surprising finding is that a switch cost is still evident when there is plenty of time to prepare for a switch—so-called *residual switch costs*. Such residual costs were thought to reflect a limitation in the cognitive system that prevents it fully preparing for a task switch in advance of the stimulus. However, De Jong (2000) constructed CDFs of task-switching performance to address whether there was such a limitation of preparatory ability. De Jong argued that at the fastest end of the distribution, participants were fully prepared to perform the task (hence, the fast responding); at the slower end of the distribution, he argued, participants were not prepared. Thus, CDF construction allowed De Jong to investigate the switch costs on prepared and unprepared trials. He found that switch costs were all but absent at the fastest end of the distribution, and increased towards the slower end of the distribution. Based on this finding, De Jong presented a model that suggested that full task preparation can occur, but that participants fail to engage in such preparation on a proportion of trials (reflected in the slower end of the distribution).

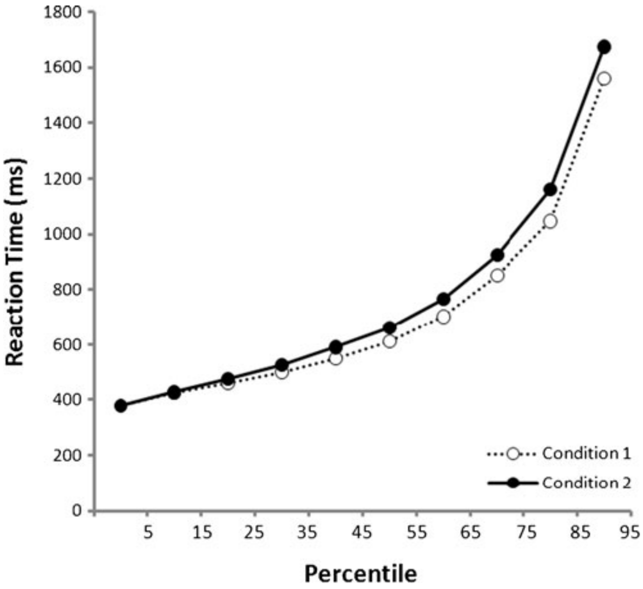


Fig. 1 Example of a cumulative distribution frequency (CDF) plot of two experimental conditions. Percentile is plotted against reaction time (in milliseconds). In this example, the first percentile is at 5% and the last is at 95%

Grange and Houghton (2011) also used the CDF method, together with reasoning similar to De Jong's (2000), to argue that inhibition in task switching can be overcome given full task preparation. Inhibition in task switching can be inferred from slower RTs when returning to a task recently performed (i.e., an ABA sequence), as compared to returning to a task not so recently performed (a CBA sequence). These n–2 repetition costs are thought to be caused by backward inhibition: In an ABA sequence, switching from Task A to Task B requires inhibition of Task A; this inhibition lingers and hampers reactivation of Task A if this is required a short time after (Mayr & Keele, 2000; see Koch, Gade, Schuch, & Philipp, 2010, for a review). It is a relatively well-established finding that increasing the time for task preparation does not reduce *n*–2 repetition costs. However, Grange and Houghton (2011) constructed CDFs of *n*–2 repetition costs and found that they *were* absent at the fastest end of the RT distribution, and increased steadily towards the slower end. This finding seems incompatible with the idea that overcoming inhibition is not possible with preparation, since the faster trials are assumed to be the most prepared (De Jong, 2000).

CDFs have also proven useful in other areas of RT research, including stimulus–response compatibility studies (Cohen, Bayer, Jaudas, & Gollwitzer, 2007; Hommel, 1996; Wascher, Schatz, Kuder, & Verleger, 2001), lexical decision tasks (Perea, Rosa, & Gómez, 2005; Yap, Balota, Tse, & Besner, 2008), visual cognition (Chun & Wolfe, 1996), and memory processes (Meiran & Kessler, 2008; Nino & Rickard, 2003; Oberauer, 2005). Proctor, Miles, and Baroni (in press) recently reviewed the literature of group RT distribution analysis for the spatial Simon effect, an RT benefit for a lateralised response when the spatial position of the stimulus is congruent with the response. Analysis of RT distributions in this paradigm has proved especially fruitful (see, e.g., Cohen et al., 2007; Hommel, 1996; Wascher et al., 2001). One example of the impact that CDFs have had in this literature is from De Jong, Liang, and Lauber (1994), who plotted CDFs of the spatial Simon effect and discovered that the effect was largest at the fastest quantile, reducing in magnitude towards the slower end of the distribution. De Jong et al. used this evidence to support their hypothesis that the Simon effect is caused by automatic activation of spatially congruent responses triggered by stimulus onset; this activation dissipates rapidly, thus only affecting the faster end of the RT distribution. Such hypotheses are impervious to mean-RT analysis alone (see Zhang & Kornblum, 1997, for further discussion).

To summarise much of the work described above, CDF analysis should be included (or at least explored) for any of the following situations:

(1) Two (or more) conditions differ with respect to an early, fast-acting process, whose effect dissipates rapidly or can be overcome with more preparation time. In this case, the effect should be largest at faster RTs (e.g., De Jong et al., 1994).

(2) Two (or more) conditions differ with respect to a process whose effects may be reduced or overcome on "fully prepared" trials, or that needs a certain amount of time to have its maximum impact on response processes. In this case, the effect should be largest at slower RTs (e.g., Grange & Houghton, 2011).

(3) Two conditions are oppositely affected by two processes with different time courses, an early and a late process. In this case, the two CDF curves would show a crossover interaction. This might occur, for instance, if different strategies were being used in the two conditions. We have seen no example of this hypothetical case, but note that, given the crossover, the "main effect" of the condition means could well be nonsignificant. Hence, null results can be further explored using CDFs. Even a demonstration that an effect is indeed null over the whole RT distribution would strengthen any attempt to place theoretical weight on its absence.

Means or medians? At the individual participant level in RT studies, investigators almost always generate condition scores using either the median or the mean of the participant's RT distribution for that condition. The choice is an important one, because the distributions of individual participant RTs are not symmetrical (i.e., mean and median do not coincide), and it is these scores that are entered into the statistical analysis. Although we have not attempted to quantify it, our impression is that the mean (usually of trimmed data) is the far more popular measure. This contrasts with conventional CDFs, which are computed on percentiles at the participant level. Percentiles, of course, are a "median-type" measure (the median is the 50th percentile, the 1st decile is the median of the first 20% of scores, etc.). Thus, if an investigator reports statistical analyses based on participant means, but then additionally analyses percentile-based CDFs for the same data, the investigator is performing related analyses using two different measures of central tendency. Some investigators may wish to avoid this mixing of measures, while still using condition means at the participant level. For this reason, CDF-XL offers an additional form of CDF analysis based on the means of rank-ordered subsets ("bins") of the RT data, rather than percentiles. This analysis stands in the same relation to the use of participant means in global statistics as do percentile CDFs to the use of the median. This relationship is described in more detail below.

To reduce redundancy in the description of these two analyses, we will use the general term *partition* to refer to any way of dividing up the data. In the context of the percentile analysis, *partition* should be taken to refer to division by percentiles, while for the bin-means analysis, it refers to the data subset from which the bin mean is calculated.

Producing CDFs Despite the greater information yield that analysis of RT distributions can produce from experiments, the use of CDFs (or other distributional analyses) remains relatively rare in published data. We believe that one major reason for this is that producing such an analysis, even just for exploratory purposes, is laborious and time consuming. We know of no experimental software or data analysis package that automates the generation of CDFs from experimental data in more or less their raw form. Even using market-leading spreadsheet software such as Microsoft Excel, the process of analysing data from a single experiment might take a number of hours. To see this, consider that to produce CDFs, the experimental data must first be separated into data from individual participants, which then must be sorted (for each participant) by condition, and the trials must then be rank ordered by RTs within each condition. The score for each quantile or data partition (possibly as many as 20 per condition) must be derived in some manner from the rank ordering of each condition. Given that the number of data points will usually vary from condition to condition (and between participants) due to prior filtering of errors or other events leading to absent data, this process resists simple automation. Finally the derived scores must be appropriately tabulated for entry into a statistical analysis such as ANOVA, and the grand means by partitions must be calculated for the generation of the CDF plots. If the investigator wishes to change the number of partitions, then most of the process has to be repeated for every participant.

Aims of CDF-XL CDF-XL aims to automate the entire process described above (including the generation of CDF plots) in Excel. In brief, the user prepares the input data in a simple three-column format on an Excel spreadsheet and then clicks on a button to run the analysis. As discussed above, CDF-XL produces analyses based on two types of "partition analysis," percentiles (the "median" RT dividing one partition from the next in the ranking) and partition means (the mean of all data within each partition). The user merely specifies the number of partitions required for the analysis, and the analysis can be rerun as often as desired with different numbers of partitions with no additional effort. We hope that the simplicity and accessibility of CDF-XL will encourage investigators to explore distributional analyses. Furthermore, by implementing the program in Excel (see below), we hope to maximise its availability to students, so that it might be used, for instance, in undergraduate projects or in classes on data analysis.

CDF-XL: description of the program

Implementation

The program is implemented in Microsoft Excel 2007, using Visual Basic for Applications (VBA), and will run in Excel under any compatible version of the Windows operating system. The implementation is also backwards compatible with Excel 2003 (a version compiled for Excel 2003 is also available). On Macs, we have tested the program under Windows emulation (running Windows XP Professional) and have found it to run without problems. We have not so far tested CDF-XL under any version of Apple Mac OS. It should be noted that support for VBA was removed from MS Office 2008 for Mac, and hence, CDF-XL will not run in this version (though it may function with Office 2004 for Mac). At the time of writing [May, 2011] it is reported that VBA support will return in the projected 2011 version of Excel for Mac.

The main reason we chose to use Excel (rather than a statistical language, such as R or MATLAB) is that, in our experience, all experimental psychologists have access to and experience with it, and they routinely use it at some point in the data analysis process. In U.K. universities, it is usually provided to all faculty and students as part of the Microsoft Office suite. This familiarity and availability means that CDF-XL can be used immediately by most researchers (including undergraduate students) and that the format of the results can be easily understood and manipulated for further processing (e.g., statistical analysis), without any need for special training or extensive tutorials. Apart from the addition of CDF-XL, Excel's spreadsheet functionality remains unchanged and may be used in the usual manner.

Start-up

CDF-XL is accessed as an Excel file with the VBA program embedded and is started by double-clicking the file icon or name in a file list. It will open as any other Excel file, but users of Excel 2007 should be aware that the file has the extension xlsm, which denotes a macro-enabled workbook. When it opens, this causes the following warning (apparently obligatory) to appear just above the worksheet, on the left: "Security Warning: Some active content has been disabled."

The user must click the box labelled "Options…" (at the end of the warning) and then select the option "Enable this content." If this is not done, Excel will still function, but CDF-XL will not be accessible.

On initial start-up, the workbook will open with two worksheets, labelled "Input Data" and "Source Data." The Input Data sheet will have the focus, *and it is where the data to be entered into the CDF analysis must appear*. Use of the Source Data worksheet is optional, but it can be used to hold a larger set of data from which subsets may be passed to the Input Data sheet for successive analyses. Use of the Source Data worksheet is described later ("Data input and formatting" section).

Input data

This worksheet appears with three column headers in the first row for Columns A, B, and C (Fig. 2). Though they are not visible in Fig. 2, instructions are also provided regarding the correct formatting of the input data. Left-aligned with Column E is a simple interface ("control panel") in which the user sets the number of partitions and starts the analysis by clicking on the button labelled "Run Analysis" (Fig. 2). When CDF-XL is first started, the data columns will be empty. If the Run button is clicked with no data present, an error message appears, alerting the user.

Data input and formatting Columns A–C must contain the data to be analysed. The data must begin on Row 2, and each row represents one data point (trial). It is formatted as follows (Fig. 2):

> *Column A (Subject):* This column must contain the labels that uniquely identify each participant. Typically they are integers, but strings containing letters are also acceptable. A participant label (number) must appear on every row (data point). It is not necessary for all of the data from a given participant to be in a contiguous block of rows, because the program will reorder the data to achieve this. Hence, data from the same participant but from, say, different experimental sessions can be added piecemeal. It is not necessary to specify the number of participants, as the program will count them.
>
> *Column B (Condition):* This column must contain the labels of the "conditions" to be compared (e.g., they might be the different levels of a factor). Every data point must have a label. The program will take the number of different labels it finds in this column to be the number of conditions. The CDFs for each condition are computed separately, and the data are aggregated on the basis of the condition labels. The data need not be presorted into conditions by the user.
>
> *Column C (RT):* This column contains the reaction time for each trial. The CDF analysis is based on the rank ordering of these RTs for each participant found in Column A and for each condition found in Column B.

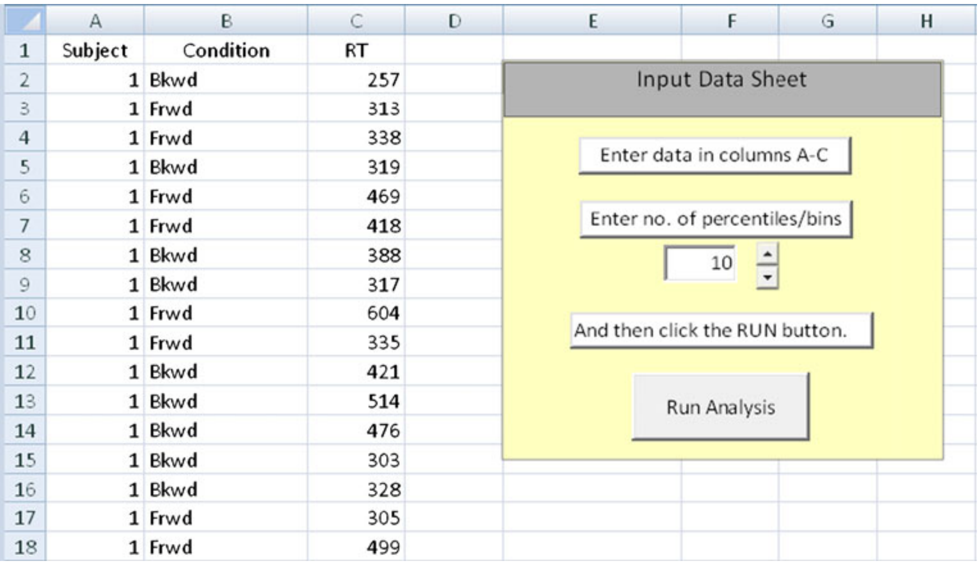
Within-group (repeated measures) comparisons

The data format described above permits within-group (repeated measures) designs to be analysed with no additional "preprocessing" of the data, as long as the comparison is between levels of the same factor (which, in all the experimental software we know of, will appear in the same output column). If the user has multifactorial data and intends to make a number of such analyses (looking at different factors), then the Source Data worksheet may be used to hold all the data, and single-factor data sets can be passed to the Input Data sheet using a custom dialog box (described below).

If the user wishes to collapse over levels that are separately represented in their raw data, then a relabelling of the conditions must be performed prior to using CDF-XL. This is easily achieved within Excel itself by, for instance, substituting the labels of all the "to be collapsed" levels with the same label.

Between-group comparisons

The same data format can also be used for between-group designs, but some relabelling of conditions may be



	A	B	C
1	Subject	Condition	RT
2	1	Bkwd	257
3	1	Frwd	313
4	1	Frwd	338
5	1	Bkwd	319
6	1	Frwd	469
7	1	Frwd	418
8	1	Bkwd	388
9	1	Bkwd	317
10	1	Frwd	604
11	1	Frwd	335
12	1	Bkwd	421
13	1	Bkwd	514
14	1	Bkwd	476
15	1	Bkwd	303
16	1	Bkwd	328
17	1	Frwd	305
18	1	Frwd	499

D
E
F
G
H
Input Data Sheet
Enter data in columns A-C
Enter no. of percentiles/bins
10
And then click the RUN button.
Run Analysis

Fig. 2 View of the top left area of the Input Data worksheet, here shown with data added to Columns A–C. When CDF-XL is started for the first time, only the column headers are present (along with the "control panel"). Each row represents one data point (e.g., trial). Column A holds subject labels, Column B, condition (level) labels, and Column C, reaction times. Once data have been entered, the user sets the number of "partitions" (percentiles or bins) in the control panel and then clicks the "Run Analysis" button

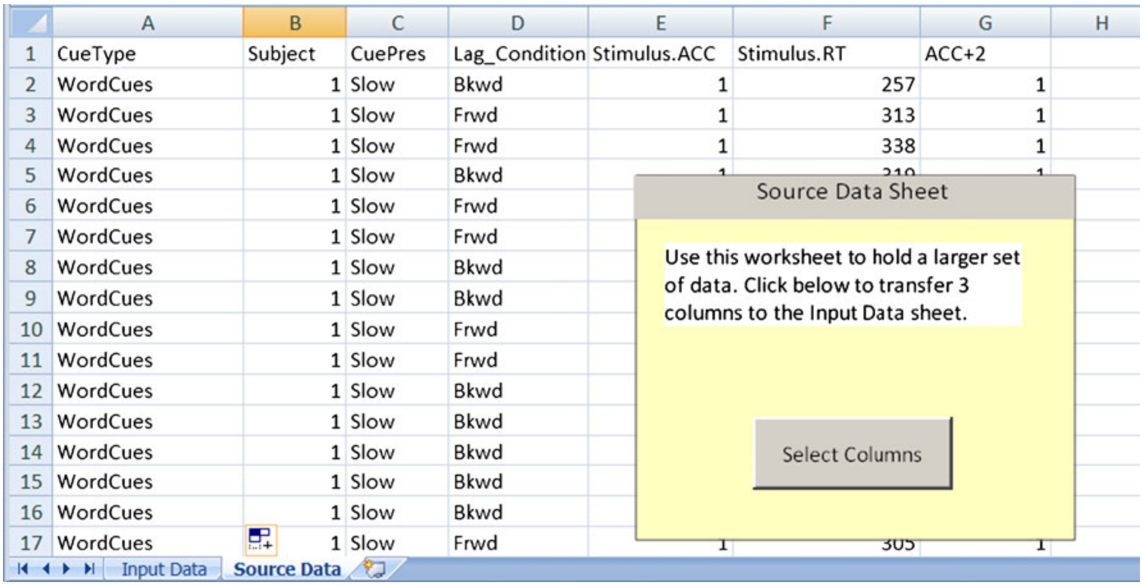
necessary. In the simplest case, in which the same (single) condition is compared between groups (e.g., a special population vs. controls), each row of the Condition column of the Input Data sheet must contain a label identifying which group each participant belongs to. There can be as many groups as required, and participants may be entered in any order with respect to group.

If more than one condition is to be compared between groups, then the limitation due to there being only one Condition column may be overcome in at least two ways. The first is to simply iterate over the process for a single between-group condition, and then combine the outputs, by copying and pasting them onto a separate spreadsheet. An alternative is to relabel the conditions, such that each trial is labelled by both group and condition in the same column. For instance, to compare two conditions, c1 and c2, between two groups, g1 and g2, would require 2 × 2 = 4 "condition" labels, which for instance could have the format g1c1, g1c2, g2c1, g2c2. Once these new labels are generated (which is easily done within Excel), the analysis can be executed as usual.

The Source Data worksheet: multiple analyses on the same data set

If the user plans to perform a number of analyses on multifactorial data, use of the Source Data worksheet (Fig. 3) removes the need to repeatedly load (formatted) data into the Input Data worksheet from an external source. It can also be used for just a single analysis if the user does not want to delete irrelevant columns from their data file by hand. When first selected, the Source Data worksheet is empty, except for the control panel containing brief instructions and a single command button. First, the user must load the data into the Source Data worksheet in whatever format they are in (though the first row of the data should contain column headers, as in Fig. 3). Then, clicking on the command button "Select Columns" (in the Source Data control panel) brings up a dialog box (Fig. 4), which allows the user to select any three columns from the Source Data (corresponding to the Subject, Condition, and RT columns, as described above) and then copies them to the appropriate columns of the Input Data worksheet.

When the dialog opens, any column headers found in Row 1 of the Source Data worksheet are shown in a labelled list box (Fig. 4), and the user must select three of these in turn, assigning them with the Subject, Conditions, and RT buttons (in any order). This is done by simply highlighting (clicking on) a column header in the list box and then clicking one of the three buttons to the right. For instance, clicking the top button will designate the currently selected column header as the Subject column, clicking the middle button will designate it as the Condition column, and so forth (see Fig. 4 for an example). Once the three column headers have been selected (the selected items appear in the text boxes to the right of the corresponding button), the user clicks the button marked "Extract Selected Columns," and the contents of the three selected columns are copied to the appropriate columns of the Input Data worksheet. From this point, the analysis proceeds as

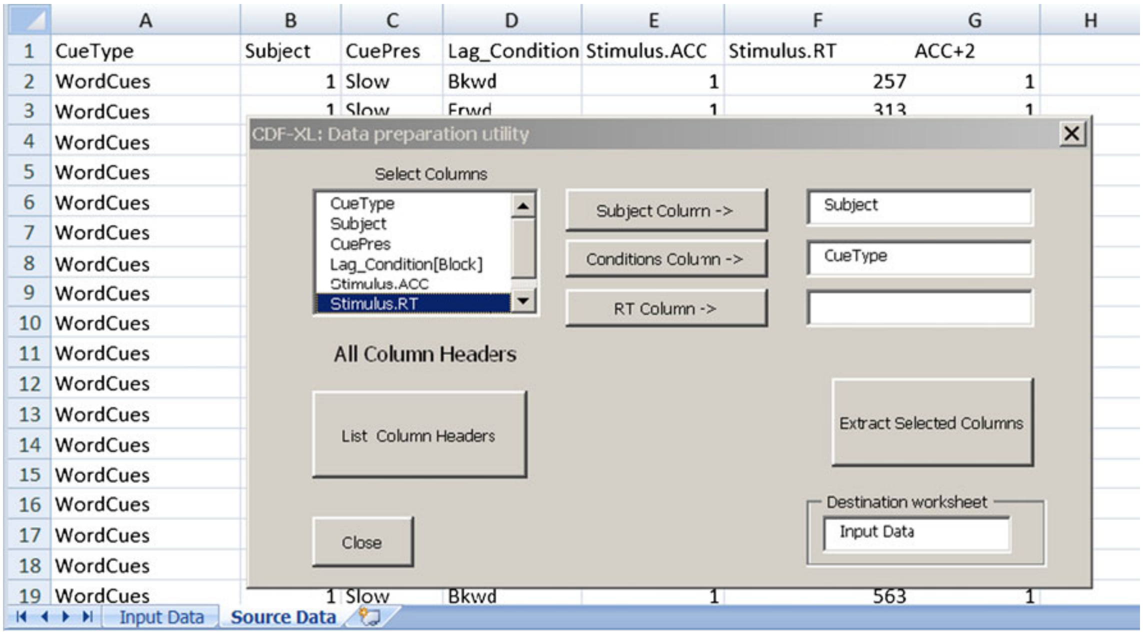


	A	B	C	D	E	F	G	H
1	CueType	Subject	CuePres	Lag_Condition	Stimulus.ACC	Stimulus.RT	ACC+2	
2	WordCues	1	Slow	Bkwd	1	257	1	
3	WordCues	1	Slow	Frwd	1	313	1	
4	WordCues	1	Slow	Frwd	1	338	1	
5	WordCues	1	Slow	Bkwd	1	310	1	
6	WordCues	1	Slow	Frwd				
7	WordCues	1	Slow	Frwd				
8	WordCues	1	Slow	Bkwd				
9	WordCues	1	Slow	Bkwd				
10	WordCues	1	Slow	Frwd				
11	WordCues	1	Slow	Frwd				
12	WordCues	1	Slow	Bkwd				
13	WordCues	1	Slow	Bkwd				
14	WordCues	1	Slow	Bkwd				
15	WordCues	1	Slow	Bkwd				
16	WordCues	1	Slow	Bkwd				
17	WordCues	1	Slow	Frwd	1	305	1	

Source Data Sheet

Use this worksheet to hold a larger set of data. Click below to transfer 3 columns to the Input Data sheet.

Fig. 3 View of the Source Data worksheet with unformatted data added. The data contain seven columns, including three "factor" columns. The participant labels and reaction times are in arbitrary columns. The user clicks the "Select Columns" button to start a utility for copying the relevant columns to the Input Data worksheet



	A	B	C	D	E	F	G	H
1	CueType	Subject	CuePres	Lag_Condition	Stimulus.ACC	Stimulus.RT	ACC+2	
2	WordCues	1	Slow	Bkwd	1	257	1	
3	WordCues	1	Slow	Frwd	1	313	1	
4	WordCues							
5	WordCues							
6	WordCues							
7	WordCues							
8	WordCues							
9	WordCues							
10	WordCues							
11	WordCues							
12	WordCues							
13	WordCues							
14	WordCues							
15	WordCues							
16	WordCues							
17	WordCues							
18	WordCues							
19	WordCues	1	Slow	Bkwd	1	563	1	

CDF-XL: Data preparation utility

Select Columns

CueType
Subject
CuePres
Lag_Condition[Block]
Stimulus.ACC
Stimulus.RT

Subject Column -> Subject

Conditions Column -> CueType

RT Column ->

All Column Headers

Destination worksheet: Input Data

Fig. 4 Use of the dialog box to copy data from the Source Data worksheet to the Input Data sheet. The Source Data column headers appear in the list box (upper left of dialog). The user must identify the three columns containing the data needed for the analysis, as follows: (i) Select (highlight) a column header in the list box with a mouse click. (ii) Click one of the three buttons to the right, depending on what the selected column represents. (iii) The selected header appears in the text box horizontally aligned with the button. In the figure, the user has already identified the columns containing the Subject labels ("Subject") and the Conditions ("CueType"), and has highlighted the header of the RT column in the list box ("Stimulus.RT"). The user now clicks the "RT Column – >" button to copy this item to the associated (empty) text box. Finally, clicking the button marked "Extract Selected Columns" copies the contents (minus the headers) of the specified columns to their appropriate locations on the Input Data worksheet. The user may return to this dialog and choose another set for analysis by selecting a different Conditions column

described below. To run a further analysis involving a different factor, the user simply returns to the Source Data worksheet and repeats the above process, selecting a different column (via its header) as the Condition column.

Running the analysis

Once the data have been entered into the Input Data worksheet, CDF-XL is run by clicking the "RUN CDF" button, which appears on the Input Data worksheet. The results of the analysis then appear shortly on a number of additional worksheets that CDF-XL creates. The labelling of these sheets and the information they contain are listed below, under Results.

Data analysis

Method

CDF:XL automatically provides two forms of partition-based CDF analysis, Percentiles (PC), and Partition (bin) Means (PM). For both analyses, the input data are first separated out by participant, and then, for each participant, rank ordered by RT within each condition. This common part of the analysis is achieved as follows:

1. The data are first sorted in situ to ensure that all data pertaining to a given participant form a contiguous set of rows.
2. The unique labels in the Subject column are extracted and counted.
3. The number of conditions and their labels are identified from the Condition column.
4. The range (start and end rows) of each participant's data is identified, and the data are copied to a new worksheet. These (single-participant) worksheets are invisible to the user, and are deleted before the analysis finishes.
5. For each participant, the data are sorted by condition, and then rank ordered by RT within condition.

The PC and PM analyses then loop through the set of (invisible) single-participant worksheets. The results of the individual-participant analyses are collated into two data tables from which the grand means (condition by partition) are computed and the CDF plots generated.

PC analysis This analysis returns the percentiles of the rank-ordered RTs by condition, for each participant. If the user specifies 10 bins, the scores will represent deciles; if the user specifies 4, then quartiles; and so forth. The percentiles required are calculated from the number of partitions entered by the user. The individual-participant results are stored in a 2-D data table with one participant to each row. The number of scores returned for each participant (columns in the data table) is the product of the number of conditions by the number of percentiles. This table is returned on its own spreadsheet (see the Results below), and from this the grand means (Condition × Percentile) are generated, to produce the percentile CDFs

PM analysis In the PM analysis, the data are transformed into the mean RTs of rank-ordered RT bins (partitions). Each bin k may be thought of as the data falling between the k and k–1 quantiles, with the quantile size specified by the user. For instance, if the user specifies 10 partitions, then the first bin (k = 1) will be the mean of the data falling below the first decile, the second bin (k = 2) the mean of the data between the 1st and 2nd deciles, and so on. This computation is first performed separately for each participant, with participant means then averaged to produce the CDF curves. The bin means by participants are returned on a separate spreadsheet in a form suitable for entry into statistical analysis.

The data (for each participant and condition) are divided into "bins" of ascending RTs, and the mean RT for each bin calculated. Given a (user-specified) number of bins N_{bins}, the number of trials, N_c, found in a condition C is divided by N_{bins} to get the "bin size" for that condition, Bin_c. In the case that Bin_c is not a whole number, it is first made equal to its integer part, $\text{Bin}_c = \text{Int}(\text{Bin}_c)$, with the result that $N_{\text{bins}} \times \text{Bin}_c < N_c$. For instance, suppose a condition contains 59 trials (N_c = 59) and N_{bins} = 10. Now $\text{Int}(N_c/N_{\text{bins}}) = \text{Int}(5.9) = 5$, and $N_{\text{bins}} \times \text{Bin}_c = 10 \times 5 = 50$. So, if only this bin size were used, 9 trials (59– 50) would be unused. To address this problem, the bin size is increased by 1 for the number of bins equal to the number of trials that would otherwise be left unused. So, in the given example, with 9 trials left over, 9 of the bins will contain Bin_c + 1 = 6 trials each, and 1 will contain 5 (9 × 6 + 5 = 59).

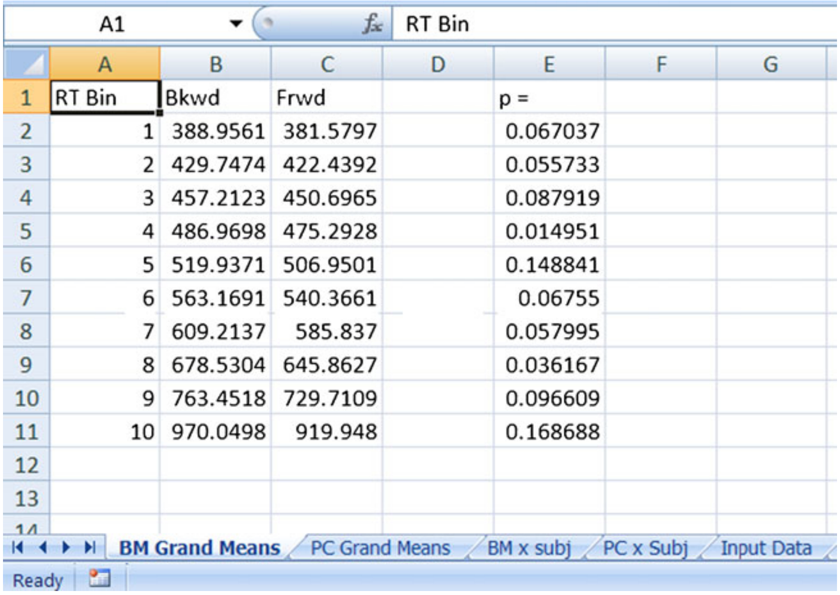
With the appropriate bin sizes calculated, CDF-XL calculates the means of the RT bins in ascending order, for all conditions for each participant. The individual-participant means are stored in a 2-D data table with the same format as for the percentile analysis. It is returned on its own worksheet, and the grand means (Condition × Bin) are computed.

Results

The results are presented on four new worksheets, two for each of the above analyses. For each analysis, one worksheet gives the individual-participant results, while the other presents the grand means along with graphs of resulting CDFs. If the analysis contains only two conditions, t tests are performed at all levels of the factor Partition (percentile or mean of RT bin).

Results by subject The two "By Subject" worksheets show the individual-participant results, one for each type of analysis. The formats of the two worksheets are identical and will be described simultaneously. In both analyses, participants appear in rows, and the partition score (percentile or bin mean) by condition in columns. The rows are "headed" (Column A) by the participant labels

Fig. 5 Snapshot of one of the results worksheets, showing the grand means (for the PM analysis) for each condition by bin. The mean RT columns (B & C) are headed by the condition labels derived from the input data. Column E shows the p value returned by a t test at each level (1–10) of the factor RT Bin



	A	B	C	D	E	F	G
1	RT Bin	Bkwd	Frwd		p =		
2	1	388.9561	381.5797		0.067037		
3	2	429.7474	422.4392		0.055733		
4	3	457.2123	450.6965		0.087919		
5	4	486.9698	475.2928		0.014951		
6	5	519.9371	506.9501		0.148841		
7	6	563.1691	540.3661		0.06755		
8	7	609.2137	585.837		0.057995		
9	8	678.5304	645.8627		0.036167		
10	9	763.4518	729.7109		0.096609		
11	10	970.0498	919.948		0.168688		
12							
13							

found in the input data. Hence, the results for individual participants can be identified and examined, if desired. The columns are ordered first by condition and then by partition within condition. The columns are labelled by the condition labels found in the data, with integers from 1 to *n* appended, where *n* = number of partitions. In this form, the data can be immediately transferred to a statistics package (e.g., SPSS) for further processing, such as analysis of variance or trend analysis.

Grand means and CDFs The grand means by condition and partition are presented on two separate worksheets, one for each type of analysis. They have the same format and will be described jointly. In the upper left of the worksheet, the grand means for each partition (percentile, bin) and condition are shown. Each row (starting from Row 2) represents one partition (percentile or RT bin). Column A shows the partition number, and the following columns show the results for each condition in the data (Fig. 5). In addition, if the data contain only two conditions, then (two-tailed) *t* tests are automatically performed at each level of the factor Partition to compare the two conditions. The results of this are shown in Column E in the form of a *p* value (Fig. 5). Although this is no replacement for a proper statistical comparison between the CDFs of the two conditions, it provides an "at a glance" indication of those points at which the bin means or percentile scores may differ reliably.

Graphs The tabular data described above may be used as input to graphical software to produce plots to the user's liking. By default, CDF-XL automatically generates Excel charts of the grand mean data, one for each of the two forms of analysis described above. The charts appear on the same worksheets as the tabulated results. Examples of charts produced by the program are shown in Fig. 6a and b, for a comparison between two conditions using 10 partitions (in both figures, results for the last partition are suppressed). In both figures, the data partition is plotted on the horizontal axis, and mean RT of the partition on the vertical. Figure 6a shows the percentile plot for the data. The 10th, 20th, 30th, and so on, percentiles by condition are computed for each participant. The curves show the sample means for percentiles 10–90 for two conditions. Figure 6b shows the "bin means" plot for the same data, in which the labelling of the horizontal axis is simply the ordinal bin number (i.e., in this case, 1 represents the first 10% of the data, 2 the second 10%, and so on). The RT data for each participant are first divided into 10 ranked bins, each containing 10% of the data (for each condition), and each bin is replaced by its mean RT. The curves show the sample means for the first 9 bins for two conditions. As can be seen, the two analyses produce very similar results. In this case, the difference between the conditions is greater at slower RTs (Grange & Houghton, 2011).

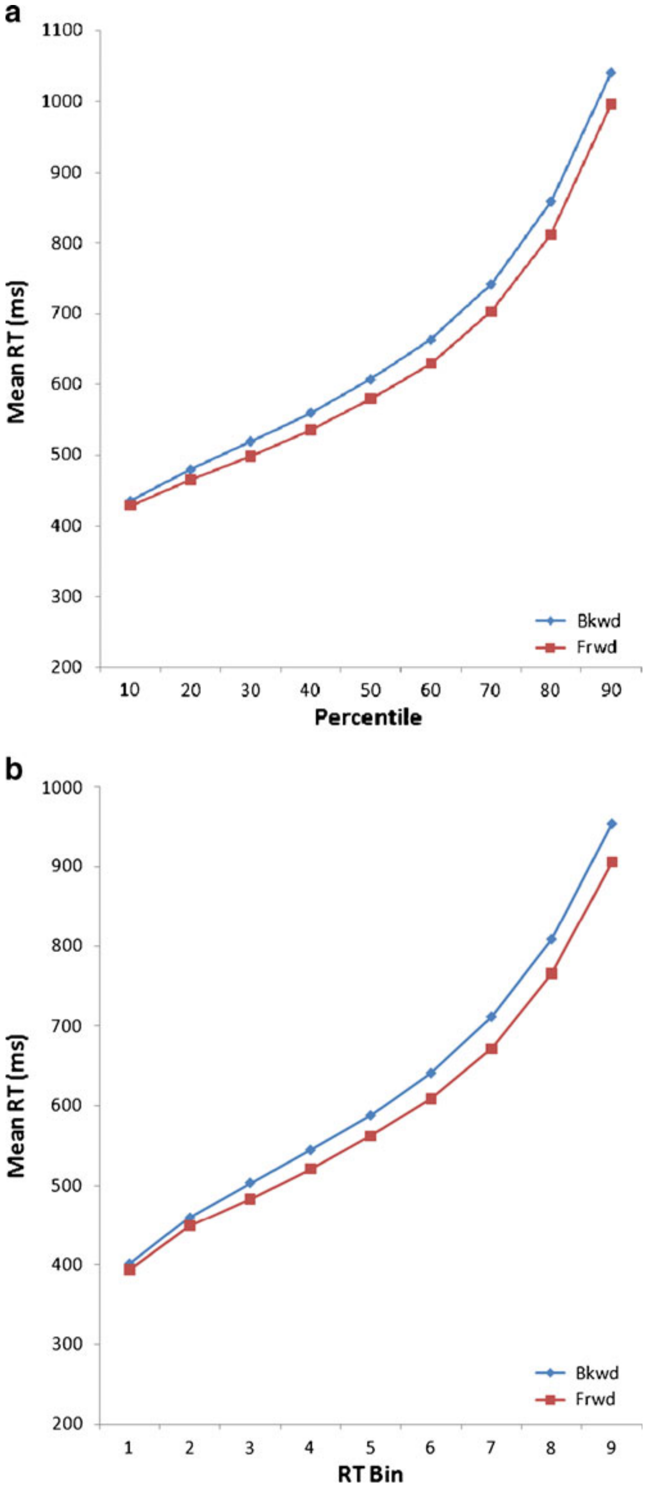


Fig. 6 Examples of plots produced automatically by CDF-XL. Both plots were generated from the same data. (**a**) CDF plot of percentiles (the 100th percentile is suppressed). The *x*-axis is labelled by percentile rather than cumulative probability (= percentile/100). (**b**) Plot of means of rank-ordered RT bins. For comparability with the percentile plot, the slowest bin is suppressed

Conclusion

The information yield from RT experiments can be increased by going beyond the use of a single measure of central tendency for each condition by supplementing it with an analysis that looks at the whole of the RT distribution, from fastest to slowest. In this article, we have demonstrated one such analysis, the cumulative distribution function. It is often the case that when using CDF analysis, one finds that an effect is constrained to be significant over only one part of the distribution, or that it changes its pattern significantly over the distribution. It is even possible, in principle, that a comparison that is null when assessed by only the central tendency may show significant differences over the CDF, for instance by "crossing over." Despite the utility of such analysis, reporting of CDFs in published work is still fairly rare, and we believe a major reason for this is the difficulty in constructing them from the typical raw-data format of an RT experiment. Popular analysis software, such as Excel, SPSS, and so forth, cannot produce them without substantial and time-consuming work on the part of the analyst. The program presented here automates such analyses with data in a format that fits many different experimental designs, and does so in Excel, a widely used spreadsheet program. We hope that in this way CDF-XL will contribute to a greater awareness and reporting of distributional analyses in published data. In addition, we hope that CDF-XL's ease of use and accessibility will be of value in teaching advanced classes on data analysis.

References

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30,* 39–78.

Cohen, A.-L., Bayer, U. C., Jaudas, A., & Gollwitzer, P. M. (2007). Self-regulatory strategy and executive control: Implementation intentions modulate task switching and Simon task performance. *Psychological Research, 72,* 12–26.

De Jong, R. (2000). An intention–activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357–376). Cambridge: MIT Press.

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus–response correspondence. *Journal of Experimental Psychology. Human Perception and Performance, 20,* 731–750. doi:10.1037/0096-1523.20.4.731

Grange, J. A., & Houghton, G. (2011). Task preparation and task inhibition: A comment on Koch, Gade, Schuch, and Philipp (2010). *Psychonomic Bulletin & Review, 74,* 481–490.

Hommel, B. (1996). S–R compatibility effects without response uncertainty. *Quarterly Journal of Experimental Psychology, 49A,* 546–571.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A.M., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin, 136*, 849–874.

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching—A review. *Psychonomic Bulletin & Review, 18,* 211–216.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology. General, 129,* 4–26.

Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology. Human Perception and Performance, 34,* 137–157.

Nino, R. S., & Rickard, T. C. (2003). Practice effects on two memory retrievals from a single cue. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29,* 373–388.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology. General, 134,* 368–387.

Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics, 67,* 301–314.

Proctor, R. W., Miles, J. D., & Baroni, G. (in press). Reaction time distribution analysis of spatial correspondence effects. *Psychonomic Bulletin & Review.*

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin, 136,* 601–626. doi:10.1037/a0019791

Wascher, E., Schatz, U., Kuder, T., & Verleger, R. (2001). Validity and boundary conditions of automatic response activation in the Simon task. *Journal of Experimental Psychology. Human Perception and Performance, 27,* 731–751.

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34,* 495–513.

Zhang, J., & Kornblum, S. (1997). Distributional analysis and De Jong, Liang, and Lauber's (1994) dual-process model of the Simon effect. *Journal of Experimental Psychology. Human Perception and Performance, 23,* 1543–1551.