

STABILITY CHECK: A PROGRAM FOR CALCULATING THE STABILITY OF BEHAVIOR

STABILITY CHECK: UN PROGRAMA DE COMPUTACIÓN PARA CALCULAR LA ESTABILIDAD CONDUCTUAL

CARLOS EDUARDO COSTA

UNIVERSIDADE ESTADUAL DE LONDRINA, BRAZIL

CARLOS RENATO XAVIER CANÇADO

WEST VIRGINIA UNIVERSITY AND UNIVERSIDADE DE SÃO PAULO, BRAZIL

Abstract

Research in behavior analysis is conducted primarily by using single-subject experimental designs, in which responding during a previous, baseline, condition serves as a control against which the experimental manipulations can be assessed. The effects of these manipulations are detected more easily when baseline behavior is stable, that is, when there is relatively little moment-to-moment variation in some aspect of behavior. The purpose of this article is to describe a computer program, *Stability Check*, that performs calculations of the stability of response rate according to two quantitative criteria. The program is simple to use, requires little hard-drive space, is free and, thus, can be useful to both researchers and teachers within behavior analysis and research methods. Detailed descriptions of the program and its use are presented after a brief summary of the importance of behavioral stability criteria in single-subject experimental designs.

Carlos Eduardo Costa, Universidade Estadual de Londrina, Brazil and Carlos Renato Xavier Cançado, West Virginia University and Universidade de São Paulo, Brazil.

We thank Dr. Michael Perone for a careful reading and discussions of a previous version of this article and for thoughtful suggestions for improving the program and its description. We would like to thank Dr. Mirari Elcoro and Dr. Carlos Aparicio for their help in translating the title and abstract to Spanish.

Correspondence concerning this article should be addressed to Carlos Eduardo Costa, Departamento de Psicologia Geral e Análise do Comportamento, Universidade Estadual de Londrina, Centro de Ciências Biológicas, Campus Universitário, Rodovia Celso Garcia Cid, Km 380, CEP: 86051-990, Caixa-Postal: 6001, Londrina, PR, Brazil. Email: caecosta@uel.br

Keywords: computer program, behavioral stability, single-subject experimental designs, behavior analysis, research methods.

Resumen

La investigación en análisis de la conducta se hace principalmente utilizando diseños experimentales de un solo sujeto, en los cuales la conducta del participante en línea base sirve como un control para estimar los efectos de las manipulaciones experimentales. Cuando la conducta en línea base es estable, es decir, cuando la variación en algún aspecto de la conducta es pequeña de un momento a otro, los efectos de las manipulaciones experimentales se pueden detectar con mayor facilidad. El propósito de este artículo es describir un programa de computación *Stability Check* que sirve para estimar la estabilidad de tasas de respuestas de acuerdo a dos criterios cuantitativos. El programa es fácil de usar, ocupa poco espacio en el disco duro, es gratuito, y puede ser útil para investigadores e instructores en los campos del análisis de la conducta y la metodología de la investigación. Después de dar un breve resumen de la importancia de los criterios de estabilidad conductual en los diseños experimentales de un solo sujeto, se hará una descripción detallada del programa y sus aplicaciones.

Palabras clave: programa de computación, estabilidad conductual, diseños experimentales de un solo sujeto, análisis de la conducta, metodología de la investigación.

The Importance of Assessing the Stability of Behavior

Research in behavior analysis is conducted primarily by using single-subject experimental designs. In these designs, initially an individual's behavior is observed and recorded during a baseline (also sometimes labeled Condition A), after which a variable of interest is manipulated in what is called a test, intervention, or treatment phase (also sometimes labeled Condition B). Comparing behavior under Conditions A and B allows the assessment of whether and, if so, how the behavior changes relative to baseline levels when other variables are manipulated. Thus, in single-subject designs, an individual's behavior during baseline serves as its own control when experimental manipulations are conducted. When repeated exposures to each condition are conducted (e.g., A and B in an ABAB design) and the changes in behavior are replicated in each condition, a higher degree of confidence is established that the manipulated variables caused the changes in behavior (cf. Baron & Perone, 1998; Barlow, Nock & Hersen, 2009; Johnston & Pennypacker, 1993; Kazdin, 1982; Matos, 1990; Perone, 1991; and Sidman, 1960/1966).

In single-subject designs, statistical control (or treatment) of the variability of some aspect of behavior that characterizes experimental designs grounded in group designs and inferential-statistical analysis is replaced by experimental control (Matos, 1990; Michael, 1974; Perone, 1999). Variability in measures of behavior is not assumed to

be intrinsic to behavior, but determined by experimental or extraneous environmental variables that can be controlled through further experimental analyses (Perone, 1999; Sidman, 1960/1966). Thus, to assess changes in behavior that occur across different phases of an experiment repeated observation and measurement of an individual's behavior need to be conducted in each phase. This can be contrasted to a single measurement of some aspect of the behavior of different individuals, as typically occurs in group-statistical designs.

The focus on experimental rather than statistical control of behavioral variability leads to an emphasis on the analysis of behavior during steady states. That is, the effects of experimental manipulations are more easily detected if baseline behavior is stable. In addition, if stability is achieved during baseline, systematic changes in behavior that occur when experimental manipulations are conducted can be more readily attributed to such manipulations, especially if a return to baseline and further replications of experimental conditions are conducted. Baron and Perone (1998) stressed the importance of the stability of behavior in single-subject designs when they stated that "stability is the foundation of single-subject research, and the evaluation of single-subject data depends on agreements that some degree of stability has been attained" (p. 50).

Behavior is considered stable when no systematic increasing or decreasing trends are observed in some measured aspect of behavior and when some regularity is observed across the observation period (Shull & Lawrence, 1998; see also Perone, 1991; and Sidman, 1960/1966). Trends or variability in the behavior under study may indicate residual effects of previous experimental conditions on current performance or, most importantly, the effects of variables that were not adequately controlled (i.e., extraneous variables) within the experimental setting. In either case, there will be problems of experimental control and in interpreting the results of a given study. It should be noted that the *absence* of systematic trends in baseline behavior is ideal, but such an absence is neither always practical or necessary to demonstrate causal relations between behavior change and the variables manipulated in an experiment. For example, if a given manipulation is expected to decrease rate of responding, an increasing trend during baseline may not be problematic, and might be acceptable, in drawing the conclusion that the manipulation caused the change in behavior.

Because behavioral stability is a fundamental aspect of single-subject designs, the methods used to assess stability are equally important in controlling the behavior of researchers. In this context, the rules for deciding whether behavior is stable, or *stability criteria*, are central not only because they can guide the decision concerning whether behavior is stable, but primarily because they affect decisions about one or another course of action in experimentation. As a consequence, using clearly defined stability criteria can reduce the influence of extraneous variables in an experiment (Johnston & Pennypacker, 1993). In addition, stability criteria also control the behavior of other researchers in evaluating and replicating published experiments. That is,

when stability criteria are clearly reported, other researchers can better assess the results and replicate the procedures.

Stability criteria used in the experimental analysis of behavior can be classified, generally, in three types (cf. Perone, 1991): (a) quantitative criteria, (b) fixed-time interval criteria (e.g., a fixed number of sessions) and (c) visual inspection of data. Quantitative stability criteria specify a given level of variability in the dependent variable. The limits of such variability can be expressed either in absolute (e.g., frequency or rate of responding across periods of observation) or relative terms (e.g., percentage of variability in rate of responding across observation periods). The next section describes a computer program, *Stability Check*, that can be used to calculate the stability of response rate based on quantitative criteria described by Joyce and Chase (1990) and Schoenfeld, Cumming and Hearst (1956).

The Program Stability Check

The Stability Check program was written in Visual Basic.NET and is compatible with Windows® XP, Vista, and 7 operating systems. When installed in a computer using these operating systems, the program requires approximately 160 KB of hard-drive space. Microsoft Framework (version 4.0) also must be installed (information about this program can be obtained through the address: <http://www.microsoft.com/downloads/>).

The language can be set to English or Brazilian Portuguese through the menu "Tools" in the program's main screen, shown in Figure 1. The main panel is divided in three other panels (labeled A, B, and C). Data from sessions or observation periods can be entered in the upper left panel (Panel A). Each value is entered individually, and can be included in the analysis by clicking the button "Add" or by pressing the "Enter" key. Figure 1 (Panel A) shows hypothetical mean response rate data for one subject in each of eight experimental sessions (for ease of analysis only whole numbers were included in the examples described in what follows, but the program allows for the inclusion of values containing decimal places).

Data that have already been typed can be saved by clicking the "Save Data" button. A Windows® dialog box is opened, allowing selection of where to save the data in the computer's hard drive or in a portable drive. Data are saved in a text file (.txt) generated automatically by the program. If data from a new session are to be included in the analysis, data from previous sessions do not have to be retyped. If saved, these data can be retrieved by clicking the "Open Data" button and then selecting the desired .txt file. When the .txt file is opened, data from all (previously included) sessions will be displayed in Panel A, as shown in Figure 1, and additional data to be included in the analysis then can be typed.

Data from a given session can be edited, removed or inserted among data from the sessions previously included in the analysis. This can be accomplished first by selecting one of the values (for example, in the data shown in Panel A of Figure 1, the rate

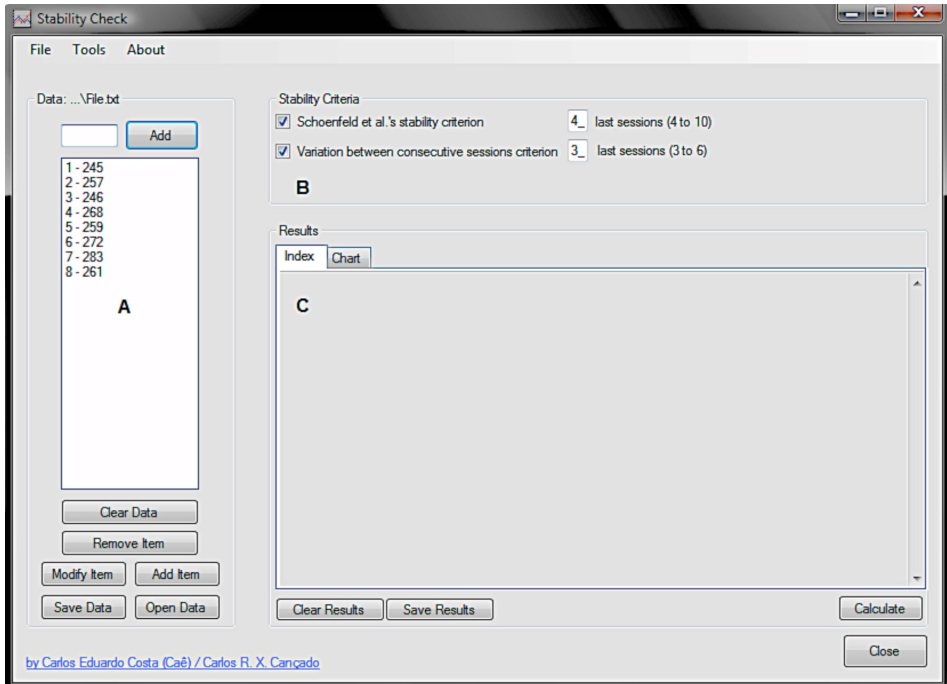


Figure 1. Main screen of the program Stability Check, displaying hypothetical data from eight experimental sessions (Panel A).

of responding of the third session, which would be 246 responses per minute) by clicking on it and by choosing to remove (by clicking the “Remove Item” button), modify (by clicking the “Modify Item” button) or add (by clicking the “Add Item” button) one value. If one value is added, it will be included after the item that was first selected (i.e., after “246”, in this example).

In the upper right panel of Figure 1 (Panel B), two options are listed for the calculation of stability. The first is the stability criterion described by Schoenfeld et al. (1956; see also Cumming & Schoenfeld, 1960), which is used extensively. Schoenfeld et al. described this criterion in a study in which data from the last six sessions of each experimental condition were included in the analysis. According to the authors,

The first seven days on any schedule are not considered in computing stability. For the next six days the mean of the first three days of the six is compared with that of the last three days; if the difference between these means is less than 5 per cent of the six days’ mean, the bird is considered to have stabilized and is shifted to the next schedule. If the difference between submeans is greater than 5 per

cent of the grand mean, another experimental day is added and similar calculations are made for that day and the five immediately preceding it. Such extensions of the experiment and calculations of stability are continued daily until the bird reaches the afore-mentioned 5 per cent criterion (p. 567; italics added).

The equation for calculating stability and described in the above quote is as follows. Considering data (D) from six sessions (D_1 through D_6),

$$\left| \frac{\left[\left(\frac{D_1 + D_2 + D_3}{3} \right) \right] - \left[\left(\frac{D_4 + D_5 + D_6}{3} \right) \right]}{\left(\frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6}{6} \right)} \right| \times 100 \quad \text{Equation 1}$$

In case stability is not achieved according to the criterion, data from another session are added (D_7) and data from the first session (D_1) are excluded from the analysis. Thus, Equation 1 would be written,

$$\left| \frac{\left[\left(\frac{D_2 + D_3 + D_4}{3} \right) \right] - \left[\left(\frac{D_5 + D_6 + D_7}{3} \right) \right]}{\left(\frac{D_2 + D_3 + D_4 + D_5 + D_6 + D_7}{6} \right)} \right| \times 100 \quad \text{Equation 2}$$

and these changes in calculations would be conducted until the stability criterion as previously described (Schoenfeld, et al., 1956) is achieved.

This process can be illustrated with an example using the program Stability Check. To use this criterion to assess stability, the option *Schoenfeld et al.'s stability criterion* should be selected, and the number of sessions to be considered for analysis should be typed in the corresponding text box located in the right portion of Panel B. For example, if the number 6 is entered in this text box, stability will be calculated for the last six sessions of the experiment included in the analysis (i.e., the last six values listed under "Data" in Panel A of Figure 1). The program allows the selection of only even numbers between 4 and 10 as the number of sessions to be included in the analysis. Considering the hypothetical data of Figure 1, the last six sessions would correspond to the values $D_3 = 246$; $D_4 = 268$; $D_5 = 259$; $D_6 = 272$; $D_7 = 283$ and $D_8 = 261$. Thus,

$$\left| \frac{\left[\left(\frac{246 + 268 + 259}{3} \right) \right] - \left[\left(\frac{272 + 283 + 261}{3} \right) \right]}{\left(\frac{246 + 268 + 259 + 272 + 283 + 261}{6} \right)} \right| \times 100 = \mathbf{5.41\%}$$

Equation 3

This result, shown in Figure 2 in the “Results” box, in the “Index” tab (Panel C), indicates that the variation in the data between the first and the second block of three sessions is 5.41% of the mean of the six sessions included in the analysis.

The level of variability that is acceptable depends on the goals of the research project, the experimental conditions in effect and, primarily, the degree of experimental control that is achievable under the conditions of the experiment. In addition to identifying the acceptable level of variation in a given experiment, the researcher decides the number of sessions that are included in the analysis. Stability Check allows calculations of stability that include data from the last four, six, eight or ten sessions.

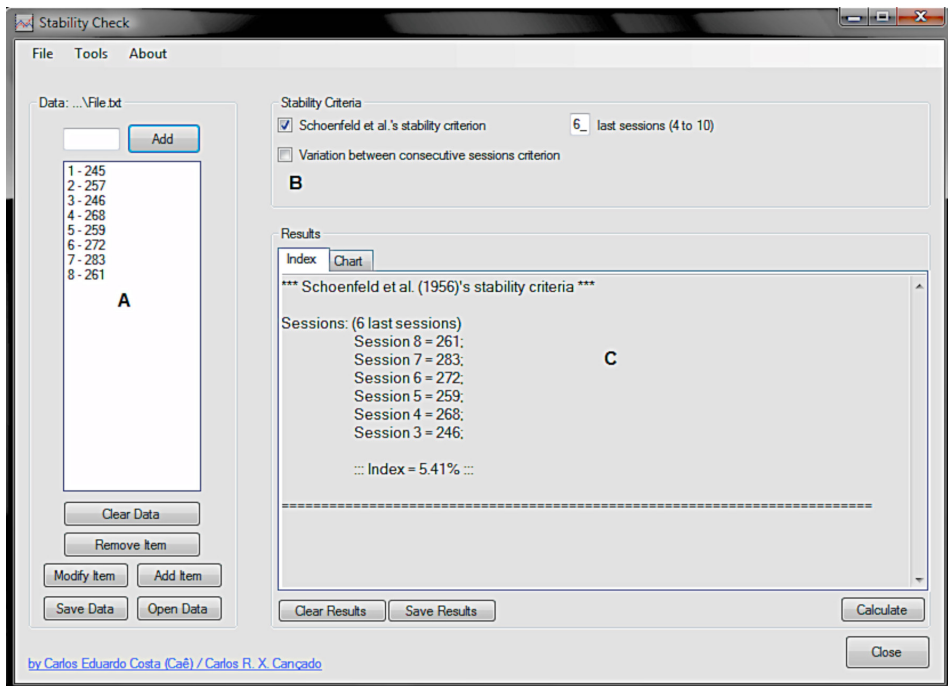


Figure 2. Main screen of the program Stability Check, with hypothetical data from eight experimental sessions (Panel A) and the results of calculations of stability (Panel C) according to the criterion described by Schoenfeld et al. (1956). See text for details.

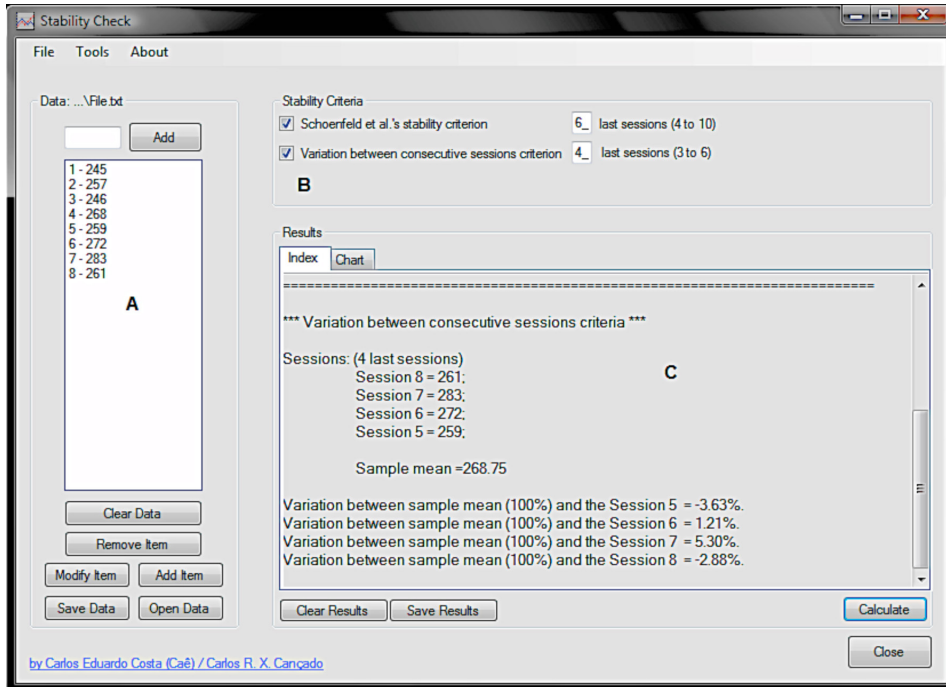


Figure 3. Main screen of the program Stability Check, with hypothetical data from eight experimental sessions (Panel A) and the results of calculations of stability according to the criterion of variation between consecutive sessions (Panel C). See text for details.

This restriction on the maximum number of sessions to be included in calculating stability was programmed because when means are used to calculate stability, the higher the number of sessions included in the analysis, the higher the probability of variability in the data being hidden, or averaged out, by the calculation (Baron & Perone, 1998; Perone, 1991).

If the *Variation between consecutive sessions criterion* (a criterion used, e.g., by Joyce & Chase, 1990; see Panel B in Figures 1 and 2) is selected, the program will calculate the variability of each session relative to the mean of all sessions included in the analysis. For example, if “4” is typed in the corresponding text box, calculations of stability will be performed considering the percentage variability of each of the last four sessions relative to the mean of these four sessions.

The results of these calculations are shown in Figure 3 (Panel C). The mean rate of responding in Sessions 5, 6, 7 and 8 was 268.75. Considering the mean as 100%, one could calculate the percentage difference between this and those values obtained in

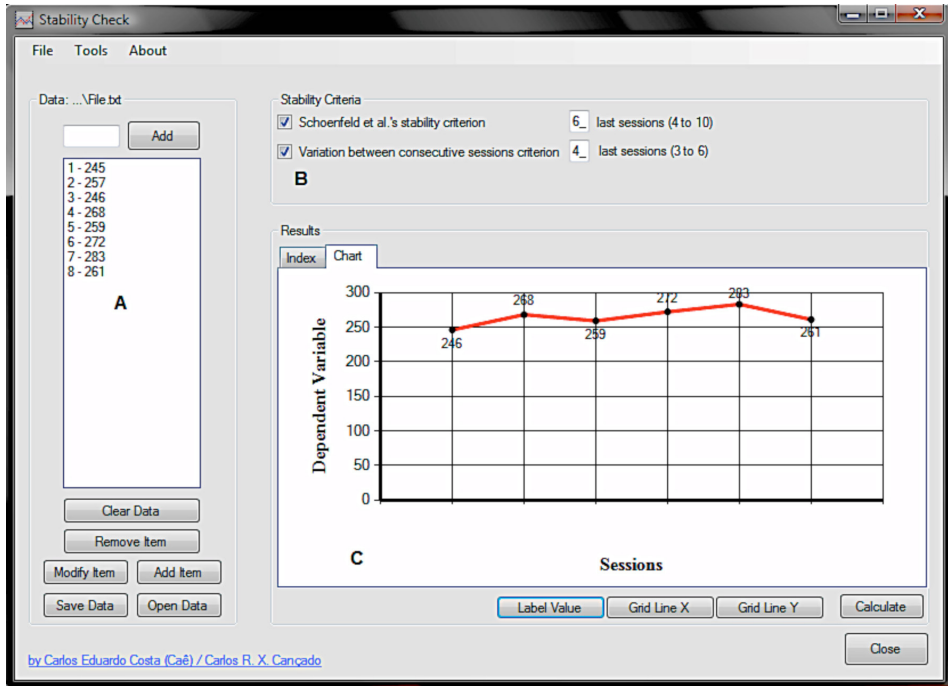


Figure 4. Main screen of the program Stability Check, with hypothetical data from eight experimental sessions (Panel A). The data for the last six sessions are shown in graphic format (Panel C), allowing visual inspection of the data used in calculating stability according to both criteria. See text for details.

each session. As shown in Figure 3, the mean rate of responding in Session 5 was 3.36% lower (note the negative sign associated with this value) than the mean of the four sessions included in the analysis. That is, if 3.36% of 268.75 (i.e., 9.75) were added to 259 (the value corresponding to Session 5) the value corresponding to the mean of the four sessions (i.e., 100% or 268.75) would be obtained. The accepted level of variability in the data will depend on those variables mentioned above.

The results can be saved in a separate text file (.txt) from those containing the data of each session, as shown in Panel A (Figures 1, 2 and 3). This can be done by clicking the "Save Results" button at the bottom left portion of the "Results" panel (Panel C), and by selecting where to save the results.

After calculating stability according to the previously described criteria, the data can be inspected visually in graphic format (in the "Results" panel, under the "Chart" tab; it should be noted that before stability is calculated, both Y- and X-axes are displayed, but no data are shown in graphic format). This graph allows detection of

trends in the data, even when stability is achieved according to the quantitative criteria described previously. In Figure 4, the data for the last six sessions (of the eight hypothetical sessions used in Figures 1 through 3) are shown graphically. This graph, with gridlines and labels for each data point, is the default generated by the program. The graphic display of data can be altered by excluding labels and grid lines by clicking on, respectively, the "Label Value" and "Grid line X" and "Grid line Y" buttons. This latter function of the program also can be used in teaching about stability of behavior by allowing the student to relate easily the two quantitative criteria previously described to the criterion of visual inspection.

Finally, it should be noted that how stringent a stability criterion is depends on how behavior occurs during baseline (e.g., if baseline behavior is high or low in rate). This factor also should be considered in deciding which stability criterion to use. Quantitative stability criteria according to which variability is assessed in relative terms (i.e., as the difference in percentage from the terminal sessions of baseline), as those described in the present article, allow greater variation when baseline response rates are high rather than low. Similarly, when variability is assessed in absolute, instead of relative differences, greater variation is allowed when response rates are low than when they are high. Thus, although Stability Check allows the calculation of stability according to two such quantitative (relative) criteria, it first should be decided if these criteria are appropriate by considering the rate at which behavior occurs in the baseline (for a detailed discussion, see Perone, 1991, p.141-144).

Conclusions

The Stability Check program can be used to calculate the stability of response rate according to two quantitative stability criteria (e.g., Joyce & Chase, 1990; Schoenfeld et al., 1956) and can be useful to both researchers and teachers using single-subject designs. The program can also be used to calculate the stability of other aspects of behavior, as long as stability is estimated according to these two stability criteria (if stability is being estimated by using a different criterion, the program can still be used if the calculations it performs are needed). Although the calculations described here can be conducted effectively by using electronic spreadsheets, using them would require all the formulas previously described to be typed before calculations were performed, increasing the probability of making mistakes. This process is automatically done when using the Stability Check program, which also is reliable, user friendly and can be downloaded at no cost through the website www.caecosta.com.br/stabilitycheck.html.

References

- Baron, A., & Perone, M. (1998). Experimental design and analysis in the laboratory study of human operant behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 45-91). New York, NY: Plenum Press.

- Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Pearson Education.
- Cumming, W. W., & Schoenfeld, W. N. (1960). Behavior stability under extended exposure to a time-correlated reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 3, 71-82. doi: 10.1901/jeab.1960.3-71
- Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Joyce, J. H., & Chase, P. N. (1990). Effects of response variability on the sensitivity of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, 54, 251-262. doi: 10.1901/jeab.1990.54-251
- Matos, M. A. (1990). Controle experimental e controle estatístico: a filosofia do caso único na pesquisa comportamental. *Ciência e Cultura*, 42, 585-592.
- Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, 7, 647-653. doi: 10.1901/jaba.1974.7-647
- Kazdin, A. E. (1982). *Single-Case Research Designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.
- Perone, M. (1991). Experimental design in the analysis of free-operant behavior. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental Analysis of Behavior, Part 1* (pp. 135-171). New York, NY: Elsevier Science.
- Perone, M. (1999). Statistical inference in behavior analysis: Experimental control is better. *The Behavior Analyst*, 22, 109-116.
- Schoenfeld, W. N., Cumming, W. W., & Hearst, E. (1956). On the classification of reinforcement schedules. *Proceedings of the National Academy of Sciences*, 42, 563-570. doi: 10.1073/pnas.42.8.563
- Shull, R. L., & Lawrence, P. (1998). Reinforcement: Schedule performance. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 95-129). New York, NY: Plenum Press.
- Sidman, M. (1966). *Tactics of scientific research: Evaluating experimental data in psychology*. Boston, MA: Authors Cooperative, Inc., Publishers. (Originally published in 1960).

Received: December 4, 2011

Final Acceptance: March 17, 2012