Chapter 11
Analysis of Intensive Categorical Longitudinal Data

Alexander von Eye and G. Anne Bogat

Intensive longitudinal data are defined as data that come from more than the usual three or four observation points in time yet from fewer than the 100 or more required for time series analysis (Walls & Schafer, 2006). Consider, for example, a clinical design with 20 repeated observations. Data from this design are hard to analyze. Unless the sample is very large, 20 observations are too many for structural modeling. For repeated measures ANOVA with polynomial decomposition, polynomials of up to the 19th order would have to be estimated (which is the easy part) and interpreted (which is the hard part). This applies accordingly to hierarchical linear models of this design. For longitudinal, P-technique factor analysis, 100 observation are needed. In brief, data that are *intensive* in the sense that more observations are made over time than usual pose specific analytic problems.

This situation is exacerbated when categorical data are analyzed. Crossing the data from 20 observation points is out of the question. Already, when only one dichotomous variable is analyzed, the number of cells of the cross-classification will be $2^{20} = 1{,}048.576$, that is, over a million cells. When multiple variables are analyzed, the situation becomes much more complex. Poisson regression models and marginal models are among the few options available for analysis (for overviews, see Agresti, 2002; Lawal, 2003). However, these options constrain the type of questions that can be asked.

In this chapter, we propose a method for the analysis of intensive categorical longitudinal data. This method is based on the well known *runs tests* (Stevens, 1939; Swed & Eisenhart, 1943; Wald & Wolfowitz, 1940). The method allows one to pursue both variable-oriented analysis, for example, log-linear modeling, and person-oriented analysis (Bergman & Magnusson, 1997; von Eye & Bergman, 2003), for example, Configural Frequency Analysis (Lienert, 1969; von Eye, 2002; von Eye & Gutiérrez-Peña, 2004).

A. von Eye (✉)
Department of Psychology
Michigan State University
East Lansing, MI 48824, USA
e-mail: voneye@msu.edu

J. Valsiner et al. (eds.), *Dynamic Process Methodology in the Social and Developmental Sciences*, 241
DOI 10.1007/978-0-387-95922-1_11, © Springer Science+Business Media LLC 2009

Runs

Consider a sequence of K scores. For this sequence, a *run* is defined as the uninterrupted sequence of $k < K$ scores that fulfill a specified condition. *Runs tests* are used to detect non-randomness in a series of scores that can be observed (e.g., in the form of serial correlation). A large number of specific definitions of runs can be used. We give three examples. First, a run can be defined as the *uninterrupted series of scores of the same value*. Consider the series 1111335222. This series contains 4 runs, including a run of $k = 4$ (value 1), a run of $k = 2$ (value 3), a run of $k = 1$ (value 5), and a run of $k = 3$ (value 2). A run can also be defined as the *uninterrupted series of scores of increasing value*. The series 1234533234 contains 2 runs. The first contains the first five scores, and the second contains the last three scores. This applies accordingly to series of decreasing values. A variant of this definition was implemented in Wallis and Moore's (1941) *runs-up-and-down* test. A third definition of a run is an *uninterrupted series of scores within a pre-specified range*. This example is of use when the reliability of machine tools is investigated or the accuracy of basketball players.

To detect non-randomness of runs, one-sample, and two-sample tests have been proposed. These tests can be reviewed in textbooks of nonparametric statistics (e.g., Bortz, Lienert, & Boehnke, 1990; Siegel, 1956), or encyclopedias (e.g., Lunneborg, 2005) and will not be described here. For the present purposes, we are not concerned with the runs tests themselves. Instead, we discuss the type of information created for runs tests, and the use of this information for the analysis of intensive categorical longitudinal data.

The Information Used to Test Runs

Series of measures have been studied from a large number of perspectives. For example, using GLM approaches, researchers have examined trends of series, changes in trends, and changes in these changes. This can be done using regression analysis and polynomial decomposition. Autocorrelation patterns and structures have been investigated using factor analysis, dynamic modeling, or, for long series, trigonometric decomposition.

In the context of longitudinal categorical data, patterns of terminal events and patterns or repeatable events have been investigated using survival analysis and event history analysis. Many other methods of analysis have been applied. For example, in the context of generalized linear modeling, marginal models and IRT models of longitudinal models have been devised. In the context of configural analysis, level, spread, autocorrelation, prediction patterns, change patterns, the symmetry of change patterns, sign patterns, and structures of first and higher differences have been investigated (von Eye, 2002; von Eye & Mun, 2007).

Runs tests use information of series that corresponds to their specific definition of runs. In addition, runs tests share the characteristic that they are based on *ordered* *sequences of scores*. That is, the order of scores must be defined. When individuals or groups are compared, the order of scores must be defined for each individual or group, but the number of scores is not necessarily required to be the same.

Runs of Scores of Same Value

As was illustrated above, the series 1111335222 contains the run sequence 4 2 1 3. To establish this sequence of 4 runs, the scale level of the scores is of no importance. All that is asked is whether adjacent scores are equal or different. Therefore, runs of scores of same value use no more than nominal scale information.

Runs of Increasing, Decreasing, or Up-and-Down Patterns of Scores

The series 1234533234 contains 2 runs of increasing scores, the first involving 5 scores (12345) and the second involving 3 scores (234). This series also contains 2 runs of decreasing scores, both involving 2 scores (53 and 32). Thus, the up-and-down pattern of scores for this series is u5, d2, d2, u3. In terms of runs of scores of equal value, this series can be described as 1 1 1 1 1 2 1 1 1. To describe the runs of increasing, decreasing, or up-and-down series of scores, no more than ordinal information is needed.

Runs of Scores within a Pre-specified Range

Consider a machine tool that is supposed to produce parts that do not deviate from a pre-specified size by more than 5μ. Suppose, this machine has produced parts with the following deviations (in μ): 3 3 2 4 2 6 6 4 4. This series contains 2 runs of tools within specification. The first run involves 5 parts, and the second involves 2 parts. These two runs are separated by a run of 2 parts that are outside the admissible range. To describe these runs, ratio scale information is used. However, ranges can also be defined for ordinal or interval level data. The type of information used for this definition of runs depends on the scale level used to specify the range.

Runs in the Analysis of Categorical Data: Algorithmic Elements

Based on the descriptions given above, the following variables describe the runs in any series of K scores:

1. *Number of runs*, K_r, with $K_r \leq K$;
2. *Length of the jth run of equal scores*, k_j, with $1 \leq k_j \leq K$ and $1 \leq j \leq K$; Depending on type of run, the following variables can also be defined:
3. *Length of jth run of ascending scores*, k_j^a, with $2k_j^a \leqslant K - 1$;
4. *Length of jth run of descending scores*, k_j^d, with $2k_j^d \leq K - 1$.
5. *Length of run within a pre-specified interval*, k_j^w, with $0 \leq k_j^w \leq K$.

Of these five variables, the first is the most important. Standard runs tests ask whether the number of runs, K_r, is smaller or larger than expected. When K_r is smaller than expected, there may be a process that prevents scores from breaking the trend. When K_r is larger than expected, there may be a process that causes overly frequent change. Length of run information is partially dependent on the number of runs. This number sets limits to both the maximum and the minimum run length. Still, the length of runs is of importance. Consider, for example, a study on the effects of psychotherapy of anancastic behavior. After establishing a baseline run pattern, the beginning of therapy can be expected to cause a run pattern that suggests improvement. This can be indicated by a pattern of decreases in the frequency of anancastic behavior occurrences over time. Therapy success can be measured using, among other indicators, length of run information.

To determine the number of runs, the following procedure is easily implemented. Let x_j be the score that was observed at time j, and K the number of observation points. Then, the *number of runs of equal scores*, r_{es}, can be calculated in two steps.

1. $\delta_j = \begin{cases} 0, & if x_j = x_{j+1} \\ 1 & else \end{cases},$

where δ_j compares the scores x_j and x_{j+1}, for $j = 1, \ldots, K - 1$, and

2. $r_{es} = \left(\sum_j \delta_j \right) + 1.$

Accordingly, the *number of runs of increasing scores*, r_{is}, can be calculated in the two steps

1. $\delta_j = \begin{cases} 0, & if x_j < x_{j+1} \\ 1 & else \end{cases}$, and

2. $r_{is} = \left(\sum_j \delta_j \right) + 1.$

The *number of runs of decreasing steps*, r_{ds}, can be calculated in the two steps

1. $\delta_j = \begin{cases} 0, & if x_j > x_{j+1} \\ 1 & else \end{cases}$, and

2. $r_{ds} = \left(\sum_j \delta_j \right) + 1.$

Finally, the *number of runs within a pre-specified interval*, r_{in}, can be calculated in the two steps

1. $\delta_j = \begin{cases} 0, & if|x_j - x| < \varepsilon \\ 1 & else \end{cases}$, where x is the pre-specified target score an the pre-specified threshold, for $j = 1, \ldots, K$, and
2. $r_{in} = \left(\sum_j \delta_j \right) + 1.$

Each of these steps can be performed using the appropriate commands in general purpose statistical software packages, for example the TRANSFORM command in SYSTAT or the COMPUTE command in SPSS. Similar operations are easily implemented in spreadsheet programs such as Lotus 1-2-3 or Excel. The result of these calculations is one runs score per case per variable. Note that runs scores can be directly compared only if the number of observation points is the same for each case. If this number varies, as is natural in training or therapy studies, the observed number of runs can be related to the maximum number of runs, $r_{max} = K$. The resulting *relative number of runs*, $r_r = r/r_{max}$, can be directly compared with other r_r scores, both intra- and inter-individually. In the following section, we illustrate the use of *number of runs*, r, in the context of categorical data analysis.

Data Examples

The following two examples use data from a study on intimate partner violence (IPV; defined here as male violence toward his female partner; cf. Bogat, Levendosky, & von Eye, 2005). Two hundred and six women were initially assessed during their last trimesters of pregnancy and then subsequently assessed yearly at the birthdays of the children with whom they had been pregnant. During pregnancy, the women had IPV experiences ranging from none to severe. In order to enroll in the study, women had to be between 18 and 40 years of age, in a romantic relationship for at least 6 weeks during pregnancy, and English speaking. Women were, on average, 25.33 years of age at recruitment and had one child $(\bar{x} = 1.4)$. Respondents were 63.7% Caucasian, 25.5% African American, 4.9% Latina, and 5.9% from other ethnic/racial backgrounds. About 40% of the women were married; 50% were single, never married; and the remainder were separated, divorced, or widowed. Their median monthly income was $1500.00.

For the examples that follow, data from 204 women were used. Two of the original 206 women were removed from the analysis because they died during the course of the study. All missing data were imputed using the hot deck method implemented in PRELIS (Jöreskog & Sörbom, 2004). At each of the 5 assessments, the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the PTSD Scale for Battered Women (Saunders, 1994), and the Severity of Violence against Women Scales (Marshall, 1992) were administered.

Example 1: Log-Linear Modeling and First Order Configural Frequency Analysis (CFA) of Runs

For the first example, we use data from the first 5 observation points (pregnancy and when the children were ages 1, 2, 3, and 4), and analyze the variables Depression (D), Posttraumatic Stress (P), and Violence Status (V). Depression was dichotomously scored as 1 = respondent exceeds cut-off for clinical-level depression, 0 = else. Posttraumatic Stress was scored as 1 = respondent exceeds cut-off for clinical-level PTSD symptoms, and 0 = else. Please note that the following analyses could have been performed using the raw scores also. Here, however, we focus on runs of clinical-level depression and posttraumatic stress symptoms. Violence Status was scored as 1 = respondent reports one or more incidents of IPV that equal or exceed threats of moderate violence in the last year, and 0 = no violence or violence below cut off.

For each of these variables, the respondent was assigned a score for runs of equal numbers. The maximum score for each variable and respondent was 5 (5 time-adjacent scores were compared for each variable; thus the maximum number of runs of scores of equal value is 5). The number of cases with 5 runs in either variable was so small that these cases were subsumed under the rubric of "4 or more" runs. The resulting three scores per respondent were crossed to form a 4 × 4 × 4 contingency table with 64 cells. This table was analyzed using standard hierarchical log-linear models and standard first order CFA. For the CFA, we used the binomial test, and the Bonferroni-protected $\alpha^* = 0.00078$.

To obtain an overview of the relationships among the three variables, a log-linear model was estimated first. Results indicate that there is no way to obtain a standard hierarchical model that is more parsimonious than the saturated model. These analyses were conducted using both SYSTAT and SPSS. Both programs produced the same overall LR-X^2. However, SYSTAT indicated that the models would not converge, even after invoking the Delta option (in SPSS, Delta had been set to zero). From these results, we conclude that, at the level of variable-oriented analysis, the data cannot be represented more parsimoniously than by the cross-tabulation itself. To explore the possibility of local associations, we then conducted a configural analysis. Results from this analysis are summarized in Table 11.1.

The overall LR-$X^2 = 221.59$ ($df = 54$; $p < 0.01$) for the CFA base model suggests that this model must be rejected, and we can expect types and antitypes to emerge. Four types and one antitype were found. The first type, constituted by Configuration 111, describes women who show only one run in all three variables. Specifically, these are women who consistently are not depressed, are not experiencing posttraumatic stress symptoms, and are not victimized during any of the observation periods. Having just one run indicates extreme stability. No change took place at all. The second type is constituted by Configuration 224. This pattern describes women who had two runs each of depression and posttraumatic stress symptoms, and four or five runs of violence. The multiple runs of violence indicate that these women frequently moved in and out of situations in which there was partner violence. The

Table 11.1 First order CFA of the Cross-classification of the Runs of Depression (D), Posttraumatic Stress Symptoms (P), and Violence Status (V)

Configuration DPV	M	$\hat{m}$	P	Type/ Antitype?
111	54	21.137	0.00000000	Type
112	8	8.348	0.54304618	
113	1	4.973	0.03966917	
114	1	1.776	0.46892307	
121	13	15.922	0.27109691	
122	1	6.288	0.01260124	
123	4	3.746	0.51688854	
124	1	1.338	0.61304950	
131	3	10.431	0.00649619	
132	3	4.120	0.40822801	
133	1	2.454	0.29492050	
134	0	0.877	0.41541607	
141	1	8.510	0.00165848	
142	4	3.361	0.43375764	
143	1	2.002	0.40404538	
144	0	0.715	0.48852438	
211	4	11.009	0.01314292	
212	2	4.348	0.18831743	
213	0	2.590	0.07376018	
214	0	0.925	0.39564853	
221	11	8.292	0.21036212	
222	8	3.275	0.01810888	
223	5	1.951	0.04751055	
224	7	0.697	0.00000793	Type
231	7	5.433	0.30225885	
232	2	2.146	0.63707170	
233	1	1.278	0.63419979	
234	0	0.457	0.63313803	
241	0	4.432	0.01132156	
242	2	1.751	0.52326573	
243	1	1.043	0.71995365	
244	0	0.372	0.68880758	
311	0	7.926	0.00030826	Antitype
312	2	3.131	0.39269809	
313	0	1.865	0.15356524	

Table 11.1 (continued)

Configuration DPV	*M*		P	Type/ Antitype?
314	0	0.666	0.51315375	
321	3	5.971	0.14987523	
322	2	2.358	0.58013826	
323	2	1.405	0.41032047	
324	0	0.502	0.60510813	
331	5	3.912	0.35375670	
332	7	1.545	0.00102893	
333	6	0.920	0.00036639	Type
334	1	0.329	0.28034608	
341	8	3.191	0.01576930	
342	0	1.260	0.28243968	
343	0	0.751	0.47130478	
344	0	0.268	0.76464584	
411	2	4.844	0.13528602	
412	2	1.913	0.57119372	
413	1	1.140	0.68436033	
414	0	0.407	0.66533698	
421	1	3.649	0.11872585	
422	0	1.441	0.23546466	
423	0	0.859	0.42302351	
424	0	0.307	0.73576574	
431	1	2.391	0.30871702	
432	1	0.944	0.61184391	
433	0	0.562	0.56935403	
434	0	0.201	0.81792613	
441	6	1.950	0.01434491	
442	3	0.770	0.04291890	
443	5	0.459	0.00011131	Type
444	0	0.164	0.84878859	

two runs in the response variables suggest that these women switched from non-caseness to caseness over the course of the observations. Only 7 women showed this pattern. This number, however, is significantly above the expected number.

This applies accordingly to the 6 women who showed 3 runs in all three variables. These women moved in and out of symptom status as well as in and out of victim status. As rare as this pattern is, it was observed more often than expected. Similarly, only 5 women displayed Pattern 443, that is, the maximum number of runs of depression and posttraumatic stress symptoms, in response to moving in and out of violent partner situations. Again, this number was small but large enough to define a type. Because the occurrence rates of the last three types were so small, replication studies are encouraged to firmly establish the existence of these types.

The sole antitype was constituted by Configuration 311. Not a single respondent showed this pattern, but about 8 had been expected to show it. Exhibiting three runs of depression in combination with stable non-caseness in posttraumatic stress and stable non-violence is less likely than one would expect for a chance event.

Example 2: Log-Linear Modeling and 2-group CFA of Depression and Posttraumatic Stress Runs

In the second data example, we examine whether experiencing IPV during pregnancy confers risk for the women's subsequent mental health. As in the first example, we first perform a log-linear analysis and, then, run a 2-group CFA in which we compare the run patterns of those women who experienced IPV above the cutoff specified above during pregnancy with those women who did not. Sixty-seven women experienced above-threshold partner violence during pregnancy, and 137 did not. As in the first example, we first attempted to identify a parsimonious log-linear model, and then to perform a 2-group CFA. The table examined for these analyses is of size 4 × 4 × 2, and is spanned by the runs for Depression (D; 4 categories), PTSD symptoms (P; 4 categories), and Violence during pregnancy (V; 2 categories).

The log-linear analyses resulted in the model of all two-way interactions, [D, P][D, V][P, V]. For this model, the LR-X^2 was 9.63 ($df = 9$, $p = 0.38$), thus indicating impeccable model-data fit. A logistic regression model in which we predicted violence status during pregnancy from the two runs patterns fit well also (Hosmer-Lemeshow $X^2 = 4.13$; $df = 7$; $p = 0.765$; Nagelkerke $R^2 = 0.33$). Considering, however, that the overall classification percentage was no better than 70.1, and that two of the three contrasts that were estimated for posttraumatic stress were not significant (Category 4 was used as the reference), we ask, where, in particular, the action in the 4 × 4 × 2 (D × P × V) table can be located. To answer this question, we perform a 2-group CFA.

For this analysis, we use the z-approximation of the binomial test and the Bonferroni-protected $\alpha^* = 0.003$. Table 11.2 displays results of the 2-group CFA.

The 2-group CFA revealed three discrimination types. The first of these is constituted by Configuration 11. These are women with just one run each of depression and posttraumatic stress symptoms. This pattern is observed significantly more often in women who did not experience above-threshold partner violence during pregnancy. The second discrimination type is constituted by Configuration 22. These women slide from non-caseness into caseness (or vice versa) and remain there, during the observation period. Significantly more women who experienced violence during pregnancy exhibit this pattern than do women who do not experience violence during pregnancy. Similarly, Configuration 24 is found more often in

Table 11.2 2-group CFA of the Cross-classification of Runs of Depression (D), Posttraumatic Stress Symptoms (P) with Violence During Pregnancy (V)

Configuration DPV	*m*	*z*	*p*	Type?
111	58			
112	2	5.792	0.000000	Discrimination Type
121	11			
122	3	0.942	0.173080	
131	2			
132	0	0.994	0.160160	
141	0			
142	1	−1.434	0.075847	
211	17			
212	11	−0.782	0.217140	
221	3			
222	8	−2.896	0.001889	Discrimination Type
231	7			
232	4	−0.256	0.399038	
241	1			
242	7	−3.358	0.000392	Discrimination Type
311	11			
312	5	0.141	0.443908	
321	7			
322	6	−1.056	0.145401	
331	3			
332	5	−1.822	0.034197	
341	0			
342	1	−1.434	0.075847	
411	8			
412	7	−1.185	0.118080	
421	6			
422	3	−0.032	0.487143	
431	3			
432	4	−1.393	0.081772	
441	0			
442	0	0.061	0.475623	

women who experience violence during pregnancy than in women with no violence. These are women who slide into depression and stay there, and have repeated bouts with posttraumatic stress symptoms. Overall, the negative signs in the table suggest that the mental health of women who experience violence during pregnancy is less stable in the following years than that of women with no violence. The relatively small number of discrimination types can be explained by the low frequencies of individual configurations.

Discussion

In this article, we presented a new method for the analysis of longitudinal categorical data. This method enriches the arsenal of options of categorical data analysis in two ways. First, the method allows one to test hypotheses concerning *intensive longitudinal data*. This type of data has, thus far, been hard to analyze. In fact, even the above-mentioned marginal and regression-type models pose problems, in particular in the analysis of multivariate data. Analyzing the runs structure of longitudinal data allows for the reduction of the number of variable categories to the extent that complete cross-classifications can be examined even for samples of moderate size. Consider the first data example, above. Before extracting the runs information, we had 5 observation points for three dichotomous variables. Completely crossed, these variables would have spanned a contingency table with $2^{15} = 32.768$ cells. Tables this large require colossal samples for proper analysis. The methods presented here enabled us to analyze all three variables for all observation points, using only the 204 cases of our sample. The only data reduction that was performed involved the aggregating of two sparsely populated variable categories. The resulting cross-tabulation had $4^3 = 64$ cells, that is, a portion of 0.00195 of the big original table, a savings of 998 per mill.

Second, using runs information for categorical data analysis offers a new set of hypotheses that can be tested—hypotheses that have, to the best of our knowledge, not been proposed or tested in a multivariate, longitudinal categorical variable context. These hypotheses concern the number of runs or, in different words, a particular aspect of stability over time. Thus, the methods proposed here are the first to allow one to study *patterns of multivariate temporal stability and change*, and this can be done both from the variable-oriented and the person-oriented perspectives.

There can be no doubt that this reduction in size comes at a price. This price involves a change in hypotheses that can be tested. For example, hypotheses concerning the association structure of the original variables cannot be simultaneously entertained, and neither can hypotheses about individual patterns of change from one category to the next. These and other standard hypotheses are replaced by hypotheses about runs and their interactions.

The runs tests discussed in the literature are notorious for possessing low power. However, they are nonparametric, and parametric alternatives have not been proposed. Therefore, there are few alternatives to runs tests. In the context of this article, the power of runs tests is of lesser importance. We focus on the information that is extracted from series of measures and used when testing hypotheses involving runs.

In this chapter, we demonstrated the usefulness and versatility of the runs concept for the analysis of intensive categorical longitudinal data. In the first of the above examples, we tested and explored hypotheses concerning the joint stability of three variables. In the second example, we tested and explored hypotheses concerning the differences in stability that can be found in two groups of respondents. In the first example, no parsimonious log-linear model could be specified. Therefore, CFA was the only option for analysis. In the second example, a log-linear model and a logistic regression model described the data well. Using 2-group CFA, we identified those patterns that carry these (local) variable relationships (the relationship between 2-group or prediction CFA and logistic regression is discussed in detail in von Eye & Bogat, 2005, and in von Eye, Mair, & Bogat, 2005). Thus, the runs approach offers a new way to assess stability and change in intensive categorical longitudinal data from both the variable-oriented and person-oriented perspectives.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, *9*, 291–319.

Bogat, G. A., Levendosky, A. A., & von Eye, A. (2005). The future of research on intimate partner violence: Person-oriented and vairable-oriented perspectives. *American Journal of Community Psychology*, *36*, 49–70.

Bortz, J., Lienert, G. A., & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.

Jöreskog, K. G., & Sörbom, D. (2004). *Lisrel 8.7 for Windows* [*computer software*]. Lincolnwood, IL: Scientific Software International, Inc.

Lawal, B. (2003). *Categorical data analysis with SAS and SPSS applications*. Mahwah, NJ: Lawrence Erlbaum.

Lienert, G. A. (1969). Die Konfigurationsfrequenzanalyse als Klassifikationsmittel in der klinischen Psychologie. In M. Irle (Ed.), *Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie 1968 in Tübingen* (pp. 244–255). Göttingen: Hogrefe.

Lunneborg, C. E. (2005). Runs test. In B. S. Everitt & D. C. Howell (Eds.). *Encyclopedia of statistics in behavioral science* (p. 1771). Chichester, UK: Wiley.

Marshall, L. L. (1992). Severity of the violence against women scales. *Journal of Family Violence*, *7*, 103–121.

Saunders, D. G. (1994). Posttraumatic stress symptom profiles of battered women: A comparison of survivors in two settings. *Violence and Victims*, *9*, 31–44.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Stevens, W. L. (1939). Distribution of groups in a sequence of alternatives. *Annals of Eugenics*, *9*, 10–17.

Swed, F. S., & Eisenhart, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *Annals of Mathematical Statistics*, *14*, 66–87.

von Eye, A. (2002). *Configural frequency analysis—Methods, models, and applications*. Mahwah, NJ: Lawrence Erlbaum.

von Eye, A., & Bergman, L. R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology*, *15*, 553–580.

von Eye, A., & Bogat, G. A. (2005). Logistic regression and prediction configural frequency analysis—A comparison. *Psychology Science*, *47*, 407–414.

von Eye, A., & Gutiérrez Peña, E. (2004). Configural frequency analysis—The search for extreme cells. *Journal of Applied Statistics*, *31*, 981–997.

von Eye, A., Mair, P., & Bogat, G. A. (2005). Prediction models for configural frequency analysis. *Psychology Science*, *47*, 342–355.

von Eye, A., & Mun, E.-Y. (2007). A note on the analysis of difference patterns—Structural zeros by design. *Psychology Science*, *49*, 14–25.

Wald, A., & Wolfowitz, J. (1940). On a test whether two alternatives are from the same population. *Annals of Mathematical Statistics*, *11*, 147–162.

Wallis, W. A., & Moore, G. H. (1941). A significance test for time series analysis. *Journal of the American Statistical Association*, *20*, 257–267.

Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.