DOCUMENT RESUME

ED 365 685 TM 020 831

AUTHOR Roberts, William L.
TITLE Programs for the Collection and Analysis of Observational Data.
PUB DATE Mar 93
NOTE 12p.; Paper presented at the Annual Meeting of the Society for Research in Child Development (New Orleans, LA, March 25-28, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Behavior Patterns; Coding; *Computer Software; *Data Collection; Foreign Countries; *Microcomputers; Observation; Reliability; *Research Methodology; *Sampling; Videotape Recordings
IDENTIFIERS *Laptop Computers

ABSTRACT

Although videotape is the preferred method for recording behavior, there are many natural settings in which it is impractical or too intrusive. The ability to collect data in such settings and in real time has been enhanced by the advent of light, powerful laptop computers. This poster (with paper) describes a set of eight programs (developed in British Columbia) for collecting and analyzing focal-individual data (i.e., events and interactions) and scan-sample data (states) on such IBM-compatible machines. These computer programs can accommodate almost any user-specified set of codes, including codes for simultaneous events. Reliabilities (percent agreement and kappa), rates, frequencies, cumulative time (time-budget data), and conditional probabilities with z-scores and exact binomial probabilities can be generated. Binomial probabilities test the null hypothesis that the conditional probability is at chance levels. General contingency across the entire set of codes can be assessed by log-linear analysis, and specific patterns of contingent behavior can be identified automatically. Four figures illustrate the discussion. (Contains 3 references.) (Author/SLD)

Programs for the collection and analysis of observational data

William L. Roberts

Cariboo College

Presented at meetings of

the Society for Research in Child Development,

New Orleans, March 1993.

Running head: Observational data

Abstract

Although videotape is the preferred method for recording behavior, there are many natural settings in which it is impractical or too intrusive. The ability to collect observational data in such settings and in real time has been enhanced by the advent of light, powerful laptop computers. This poster describes a set of 8 programs for collecting and analyzing focal-individual data (i.e., events and interactions) and scan-sample data (states) on such IBM-compatible machines. They can accommodate almost any user-specified set of codes, including codes for simultaneous events. Reliabilities (percent agreement and kappa), rates, frequencies, cumulative time (time-budget data) and conditional probabilities with z-scores and exact binomial probabilities can be generated. (Binomial probabilities test the null hypothesis that the conditional probability is at chance levels.) General contingency across the entire set of codes can be assessed by log-linear analysis; and specific patterns of contingent behavior can be identified automatically.

Because these programs were developed under funding from the Social Sciences and Humanities Research Council of Canada, **the programs and manual are available without charge for purposes of research and teaching**. Please contact William Roberts, Department of Psychology, Cariboo College, P.O. Box 3010, Kamloops, B.C. V2C 5N3, Canada.

Programs for the collection and analysis of observational data

William Roberts

Although videotape is the preferred method for recording behavior, there are many natural settings in which it is impractical or too intrusive. The ability to collect observational data in such settings and in real time has been enhanced by the advent of light, powerful, laptop computers. This poster describes a set of programs for collecting and analyzing such data on IBM-compatible machines. Because they were developed under funding from the Social Sciences and Humanities Research Council of Canada, the programs and manual are available without charge for purposes of research and teaching.

Data entry

The data-entry program (FOCAL) is designed to collect focal-individual data (Altmann, 1974) recorded in "initiator-action-target" format. Because the target's response can be entered in the following 'sentence', interactions between individuals can be coded as easily as the behavior of single individuals. FOCAL can also be used to record scan-sample data (that is, states rather than events and interactions). Users specify all codes as well as the length of the focal sample. Duration is recorded automatically. See Figure 1.

Insert Figure 1 about here

During data collection, the screen displays the previous dozen data 'sentences' as well as the one being currently entered. Errors in the current line can be corrected immediately; errors in earlier lines can be flagged for correction after the session ends. 'Prompt screens' display behavior and person codes; the current focal person, the trial number, and time elapsed in the current focal sample are also displayed. Written notes can be entered at the end of each sample.

Assessing reliability

The reliability program (KAPPA) reports kappa, the standard error of kappa, and percent agreement for the codes as a set. Percent agreement is also reported for each individual code, along with the four codes most frequently confused with it. (See Figure 2.) Because observers coding in real time will necessarily be slightly out of synchrony even when they agree about which codes to use, KAPPA allows the user to specify how much time discrepancy to tolerate. As shown in Figure 3, matching sequences and events can be marked for the two data files being compared, as an aid in training observers and understanding low reliabilities. Codes can be combined into new categories without altering the original data file.

Insert Figure 2 about here

Insert Figure 3 about here

Data analysis

TIMES reports frequency, rate, and time-budget data (cumulative time as a proportion of total time observed for continuously coded data, and frequency as a proportion of total events for scan-sample data). TIMES can handle simultaneous events.

Sequential analyses can be carried out using LAG and SEARCH. For user-designated events, LAG reports conditional probabilities, their z-scores and exact binomial probabilities (as a test of the null hypothesis that the conditional probability is at chance levels), and an index for determining if enough events have been collected (Bakeman & Gottman, 1986). If requested, LAG will also produce output for a log-linear analysis, allowing contingency to be assessed for the taxonomy as a whole. (This output is in the form of an instruction file that can be run directly by BMDP.)

Using techniques described by Bakeman and Gottman (1986) and Sackett (1979), SEARCH examines LAG output and identifies significant sequences of events. **Any number of branching sequences, each up to 30 events in length, can be identified.** Sample output is illustrated in Figure 4. SEARCH will automatically generate required LAG output. The level of alpha required for significance of the individual LAG analyses is user-specified. The significance of the sequence as a whole is also assessed. Maximum sequence length can be specified. Both LAG and SEARCH permit behavior codes to be combined or data to be pooled across subjects.

Insert Figure 4 about here

Data management

Three utility programs are also included. SCAN reads FOCAL data files and reports summary statistics (total trials, events, and time observed) and checks for common errors that arise when data are edited. Thus this program is useful for tracking the work accomplished by various members of a research team. COMPARE can be used to examine back-up and current copies of files. Discrepancies are displayed in context (i.e., with surrounding lines). EDIT is a stand-alone version of the Word-Perfect-like text processor incorporated into FOCAL. It can be used to edit data or note files.

General features

Programs can be run from menus or the DOS prompt. Multiple analyses can be run from batch files. Because data and output files are DOS Text files, they can be read by standard word processing programs and statistical packages such as BMDP.

Technical support is available from the author without charge.

References

Altmann, J. (1974). Observational study of behavior: sampling methods. Behaviour, 49, 227-267.

Bakeman, R., & Gottman, J. (1986). Observing interaction: an introduction to sequential analysis. Cambridge: The Cambridge University Press.

Sackett, G. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. Osofsky (Ed.), Handbook of Infant Development (pp. 623-649). New York: Wiley.

Figure 1: a sample data file.

```
Family  41, Home Session 1. 22 JAN 1990 at 17:06   Trial  1,  Focal person= 3, Observer= Paula
2  20  3     3.9     4.0
3  20  2     1.4     5.3
2  20  3     1.2     6.5
3  20  2     3.5    10.0
3  05  3     9.6    19.6
2  20  3     9.0    28.6
2  13  3     3.4    31.9
3  05  3     2.6    34.5
3  01  3     4.6    39.1
3  20  2     1.7    40.8
2  20  3     4.6    45.4
3  78  2     5.4    50.8
3  01  3    11.4    62.1
3  05  3    16.0    78.1
3  01  3     2.0    80.1
3  78  9     2.4    82.5
9  27  3     5.8    88.3
3  01  3    40.0   128.3
 -1
Family  41, Home Session 1. 22 JAN 1990 at 17:19   Trial  2,  Focal person= 2, Observer= Paula
2  22  3    10.3    10.3
2  22  4     8.0    18.3
2  20  4    10.0    28.3
3  20  4     5.6    33.9
2  20  3     3.4    37.3
3  20  2    10.2    47.5
2  78  3     2.4    49.9
2  20  3     5.1    54.9
2  22  4     3.4    58.3
3  20  2     3.3    61.6
2  20  3     2.5    64.1
3  20  2     2.0    66.0
2  20  3     6.1    72.1
3  20  2     1.3    73.3
2  20  3     4.6    77.9
3  20  2     2.5    80.4
2  20  3    10.2    90.6
2  22  4    20.3   110.9
2  20  4     3.6   114.5
4  22  2     7.7   122.2
 -1
```

Figure 2. Reliabilities: sample output from KAPPA

Reliabilities: Per cent agreement and kappa
Tuesday, March 16, 1993 12:40 p.m.

Data from file C:\RESEARCH\TORONTO\RELIABLE\LISA104.DAT
collected 17 NOV 1988. Coder= lisa
Data from file C:\RESEARCH\TORONTO\RELIABLE\TALLY104.DAT
collected 18 NOV 1988. Coder= tally

Reliabilities calculated over 22.6 minutes across 3 trials

Permitted error in timing between sequences: 10 seconds
Permitted error when identifying sequences: 10 seconds

Code	Agreements/ Comparisons	Proportion agreement	Most frequent disagreements: code (frequency), code (frequency)			
01	8 / 10	.800	16 (1)	20 (1)		
09	3 / 3	1.000				
10	5 / 6	.833	45 (1)			
11	0 / 0	.000				
16	1 / 2	.500	25 (1)			
20	18 / 21	.857	31 (1)	45 (1)	25 (1)	
25	28 / 33	.848	10 (1)	45 (1)	26 (1)	11 (1)
26	27 / 30	.900	10 (1)	45 (1)	25 (1)	
31	0 / 0	.000				
45	1 / 1	1.000				
60	1 / 1	1.000				
63	1 / 1	1.000				

Total agreements / total comparisons= 93 / 108
Overall proportion of agreement= .861
Kappa= .825
Standard Error of Kappa= .046
Approximate 95% confidence interval= .733 to .917

Figure 3. Matching sequences and events in two reliability files.

Case ID: LISA104. Session 2. 18 NOV 1988 at 9:54. Trial 1, Focal person 3.
Observer= lisa

03 20 04	18.6	18.6		
04 25 03	5.6	24.1	E	1
03 26 05	10.0	34.1		
03 25 04	8.4	42.5	S	1
04 09 03	29.5	72.0	S	1
03 25 04	8.1	80.1	S	1
03 09 04	3.7	83.8	S	1
03 10 04	5.1	88.9	S	1
03 25 04	5.8	94.7	S	1
03 26 04	3.7	98.4	S	1
03 25 04	20.2	118.6	S	1
03 20 03	17.1	135.7	S	1
03 25 04	4.7	140.3	S	1
03 26 04	2.0	142.4	S	1
03 25 04	6.1	148.5	S	1
03 20 04	25.5	174.0	S	1
03 26 04	1.4	175.4	S	1
03 25 04	4.7	180.1	S	1
03 26 04	6.8	186.8	S	1
03 25 04	4.3	191.2	S	1
03 20 04	9.6	200.8	S	1
03 26 04	1.6	202.4	S	1
03 25 04	4.2	206.6	S	1
03 26 04	64.4	271.0	S	1
04 63 03	5.8	276.8	S	1
03 20 04	27.5	304.3		
03 25 04	3.6	307.8		
03 26 04	6.2	314.0		
05 10 03	8.5	322.5	E	2
03 25 04	33.5	356.0	S	2
03 20 04	2.2	358.1	S	2
03 25 04	11.2	369.3	S	2
03 26 04	3.6	372.9	S	2
03 25 04	12.8	385.7	S	2
03 20 04	16.0	401.7	S	2
03 26 04	1.9	403.5	E	3
03 25 04	10.1	413.6		
03 26 05	31.2	444.8		
03 20 04	36.8	481.6	S	3
03 25 04	5.8	487.4	S	3
03 26 04	6.8	494.2	S	3
03 20 04	29.7	523.9	S	3

...
-1
Agreements / comparisons= 45 / 52. Percent Agreement= 86.5

Case ID: tally104. Session 8. 18 NOV 1988 at 9:55. Trial 3, Focal person 3.
Observer= tally

03 45 04	17.1	17.1		
04 25 03	17.6	34.8	E	1
03 26 03	3.1	37.9		
03 25 04	4.2	42.1	S	1
04 09 03	7.3	49.3	S	1
03 25 04	27.6	76.9	S	1
03 09 04	4.9	81.8	S	1
03 10 04	5.4	87.2	S	1
03 25 04	10.4	97.6	S	1
03 26 04	2.4	100.0	S	1
03 25 04	16.9	116.8	S	1
03 20 03	16.1	132.9	S	1
03 25 04	5.7	138.7	S	1
03 26 04	2.5	141.2	S	1
03 25 04	5.9	147.1	S	1
03 20 04	23.6	170.7	S	1
03 26 04	2.5	173.2	S	1
03 25 04	5.9	179.0	S	1
03 26 04	6.4	185.4	S	1
03 25 04	7.4	192.9	S	1
03 20 04	7.3	200.2	S	1
03 26 04	2.0	202.2	S	1
03 25 04	4.7	206.9	S	1
03 26 04	66.8	273.7	S	1
04 63 03	11.9	285.5	S	1
03 31 04	8.3	293.8		
03 20 04	17.5	311.4		
05 10 03	7.3	318.7	E	2
03 45 05	3.3	322.0		
03 25 04	19.6	341.6	S	2
03 20 04	10.6	352.2	S	2
03 25 04	17.8	370.0	S	2
03 26 04	4.7	374.6	S	2
03 25 04	2.8	377.5	S	2
03 20 04	24.2	401.7	S	2
03 25 04	7.4	409.1		
03 26 04	8.1	417.3	E	3
03 25 04	5.4	422.7		
03 26 04	19.7	442.4		
03 20 04	37.4	479.8	S	3
03 25 04	7.5	487.3	S	3
03 26 04	4.7	492.0	S	3
03 20 04	30.1	522.2	S	3

...
-1
Agreements / comparisons= 45 / 52. Percent Agreement= 86.5

Figure 4. Patterns of behavior: output from SEARCH

Data read from file C:\STATS\918.DAT.
Required alpha= .025
Saturday, August 22, 1992 11:33 a.m.

Sequence 1 of 8
This sequence followed the criterion 7 of 93 times (.075);
expected probability= .006, z= 8.840, binomial p= .0000

Lag	Response			frequency	lag p	z	binomial p	Stability index
0	1	50	2					
1	1	10	2	29	.312	9.317	.0000	230.7
2	1	11	2	66	.985	30.185	.0000	15.8
3	1	30	2	25	.373	7.944	.0000	251.4
4	1	20	2	19	.204	4.038	.0001	174.8
5	1	21	2	83	.976	29.309	.0000	24.7
6	1	31	2	29	.341	7.813	.0000	241.6
7	1	40	2	73	.785	22.733	.0000	181.5
8	1	41	2	88	.946	28.045	.0000	54.7

[Sequence 2 omitted]

Sequence 3 of 8
This sequence followed the criterion 13 of 93 times (.140);
expected probability= .013, z= 11.006, binomial p= .0000

Lag	Response			frequency	lag p	z	p	Stability index
0	1	50	2					
1	1	30	2	47	.505	13.526	.0000	268.7
2	1	10	2	29	.312	9.317	.0000	230.7
3	1	11	2	66	.985	30.185	.0000	15.8
4	1	20	2	22	.328	7.188	.0000	237.1
5	1	21	2	83	.976	29.309	.0000	24.7
6	1	31	2	29	.341	7.813	.0000	241.6
7	1	40	2	73	.785	22.733	.0000	181.5
8	1	41	2	88	.946	28.045	.0000	54.7

[Sequences 4 to 8 omitted.]

Notes: Under the null hypothesis, the expected probability for observing a sequence is the product of its lag 1 probabilities; this value is reported above. The _z_ score is the familiar _z_ test for a proportion. Sequence 3 diverges from Sequence 1 for lags 1 to 3, then rejoins at lag 4.