Discussion

Strengthening arguments based on scale levels?

Alexander von Eye[1] and
Wolfgang Wiedermann[2]

[1]Michigan State University
[2]University of Missouri, Columbia

E-mail to corresponding author:
voneye@msu.edu

In a recent contribution to this journal, Bergman (2017) argued that the moderate yet significant Bravais-Pearson correlation of $r = 0.43$ between aggression at age 10 and age 13 fails to convey the person-oriented message that only 27% of children remain stable in their level of aggression (as measured on a 7-point Likert scale; 1 = *very low aggression*, …, 7 = *very higher aggression*). In fact, the author shows that 11% of the respondents score below average in aggression at age 10 but above average at age 13, or vice versa. A total of 19% move from low aggression to average or from high to average. In all, Bergman (2017) argues that the conclusion that there exists moderate stability is weak at best, and cannot be used to derive person-oriented conclusions concerning the *individual development* of aggression at the beginning of puberty.

In this note, we (1) take the liberty of taking issue with the statistical methods used for this discussion, (2) re-analyze the data, and (3) conclude that Bergman's (2017) arguments are defensible regardless of the methods used for analysis.

1. The data discussed by Bergman are presented in the form of a 7 x 7 cross-classification of the aggression scores of children before and after a 3-year time interval. The scale level of Likert scales has been discussed widely in the literature. The currently most accepted argument is that, when the number of scale points is increased to at least 11, Likert scales can be treated as interval level scales, without dramatic bias or loss of information (Wu & Leung, 2017). When Likert scales have fewer than 11 scale points, they are best treated as ordinal in nature. From this perspective, it can be viewed as questionable that the Bravais-Pearson correlation coefficient was used to relate the two sets of aggression scores to each other. Although Pearson's correlation formula $r = \text{cov}(x, y) / (s_x s_y)$ (with $\text{cov}(x, y)$ being the covariance and s_x and s_y being the standard deviations of x and y) can be used as the basis to derive the phi-correlation for nominal variables and Spearman's correlation coefficient for ordinal variables, the correlation coefficient r requires at least interval-level scales. On the other hand, adopting an argument that was used in a discussion of Stevens' (1946) scales by Hand (1993) and Velleman and Wilkinson (1993), scale levels are inept at determining which statistical method be used for data analysis. From this perspective, Bergman's (2017) use of the correlation coefficient r is certainly defensible. Here, we do not ask whether the use of r is incorrect or can be defended. Instead, we ask whether exploiting the information that is inherent in scales at particular scale levels can lead to different conclusions about the data structure. We, therefore, re-analyze these data using different methods.

2. When the correlation coefficient r is used to describe the association between the aggression scores that were taken three years apart, one assumes that the 7-point Likert scales carry the information that comes with interval scales. We now re-analyze these data under the assumptions that the Likert scales operate at Stevens' (1946) (1) nominal and (2) ordinal scale levels. Table 1 displays the observed frequencies and the expected frequencies that were calculated under these two scale level assumptions. To calculate the expected cell frequencies under a nominal-scale level, we used Pearson's chi-square. To calculate the expected frequencies under an ordinal-scale level, we used a method proposed by Haberman (1974; cf. Fienberg, 1981). That is, in both models, we assume that the scale level of the Likert scales is below the interval level. Haberman's approach implies using the ranks of ordinal scales as covariates of a log-linear model. Alternative approaches (that might result in different expected cell frequencies) are summarized in Fullerton's (2009) conceptual framework for ordered logistic regression models.

Table 1.

Cross-classification of aggression scores at ages 10 and 13, for 916 children (from Bergman, 2017, p. 121); Cells display observed frequencies on top, expected frequencies under a nominal scale model in the middle, and expected frequencies under an ordinal scale model at the bottom of each cell.

Aggression at Age 10	Aggression at Age 13							
	1	2	3	4	5	6	7	Total
1	32	31	14	20	8	2	0	107
	10.15	15.89	18.11	33.17	14.60	10.16	4.91	
	30.18	30.35	16.77	22.64	5.03	1.89	0.16	
2	18	30	28	48	11	6	3	144
	13.68	21.38	24.37	44.65	19.65	13.68	6.60	
	19.65	29.44	26.13	45.28	14.67	7.66	1.17	
3	20	29	37	57	19	15	1	178
	16.91	26.43	30.12	55.19	24.29	16.91	8.16	
	18.63	31.69	34.39	58.27	20.96	11.56	2.47	
4	12	37	54	105	52	30	10	300
	28.49	44.54	50.76	93.01	40.94	28.49	13.76	
	15.54	34.89	53.83	103.65	48.13	32.61	11.31	
5	3	5	11	38	19	17	9	102
	9.69	15.14	17.26	31.62	13.92	9.69	4.68	
	2.13	6.46	14.49	32.41	19.86	16.89	9.74	
6	2	4	10	10	10	12	11	59
	5.60	8.76	9.98	18.29	8.05	5.60	2.71	
	0.80	2.77	7.67	16.95	11.64	10.57	8.62	
7	0	0	1	6	6	5	8	26
	2.47	3.86	4.40	8.06	3.55	2.47	1.19	
	0.08	0.39	1.72	4.79	4.70	5.81	8.54	
Total	87	136	155	284	125	87	42	916

The standard model of independence of the aggression scores at ages 10 and 13 comes with a Likelihood Ratio Chi-square of 218.98 ($df = 36$, $p < 0.01$). This value suggests a strong association between the aggression scores at ages 10 and 13. If the underlying assumption is correct that the scales carry no information above and beyond that carried by a nominal scale, (1) this result can be interpreted as suggesting strong cross-age stability, (2) individual residuals can be interpreted in the sense of a Configural Frequency Analysis, thus switching from a variable-oriented to a person-oriented perspective, and (3) analyses that are based on the assumption that the aggression scales operate at higher-than-nominal scale levels will not lead to different results.

If, however, the assumption is made that the scale categories are ordered, the interpretation of a strong association is inadequate. We, therefore, made the assumption that the scale categories represent ordered ranks, and re-estimated the expected cell frequencies. The Likelihood Ratio Chi-square for this model is 22.51 ($df = 25$; $p = 0.61$), thus suggesting independence between the scores from the two points in time.

3. *Conclusions*. Evidently, the results obtained under different model assumptions differ quite dramatically. Under the assumptions that the 7-point Likert scales that were used in the study on adolescent development (1) carry no more than nominal scale level information or (2) carry the interval-level information of equal intervals of scale points, one would conclude stability of aggression over time. In contrast, under the assumption that the 7-point Likert scales represent ordered categories of unspecified distances between ranks, one concludes that there is lack of stability.

Discussion

In this note, we demonstrate that, when the assumptions differ that researchers make when analyzing data, results can change dramatically. In the present example, results from the same data either suggest moderate behavioral stability over time or complete lack of stability. The questions clearly are: "which method/assumption is more appropriate," and "which result can be defended?" To be able to answer these questions, we need to know the intentions of the researchers. When global statements are intended that describe the population,

researchers need to demonstrate that their scales are properly treated as nominal – ordinal – interval level. Based on the simulation results by Wu and Leung (2017), one can be tempted to consider the 7-point Likert scales as ordinal. In this case, aggressive behavior in early adolescence is concluded to be unstable.

However, there is more to the data than broad-stroke, variable-oriented statements would suggest. Bergman (2017) shows that inferring ordinal meaning to scale points such as, for example, "below the middle rank" can lead to statements about individual development from a person-oriented perspective. In Bergman's article, these statements were based on counting instances. One could consider estimating statistical measures that set such statements in relation to expected values. Examples of such measures are the tests that are used in Configural Frequency Analysis to determine whether groups of cells constitute Types or Antitypes (see von Eye, 2002).

In sum, we distinguish three arguments that lead to a selection of methods of analysis:
1. Intentions of researchers;
2. Scale level of variables; and
3. Modeling assumptions.

The present note demonstrates that results of analysis can strongly depend on the decisions concerning these arguments. We recommend that researchers make their decisions explicit before proceeding to data analysis.

References

Bergman, L. R. (2017). Interpretation of single individual's measurements. *Journal for Person-Oriented Research*, *3*, 119 - 126. doi: 10.17505/jpor.2017.10

Fienberg, S. E. (1981). *The analysis of cross-classified data*. Cambridge, MA: The MIT Press.

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods & Research*, *38*, 306 - 347. doi: 10.1177/0049124109346162

Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics*, *30*, 589 - 600. doi: 10.2307/2529224

Hand, D. J. (1993). Comment on "Nominal, ordinal, interval, and ratio typologies are misleading". *The American Statistician, 47*, 314 - 315.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Velleman, P. F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, *47*, 65 - 72. doi: 10.2307/2684788

von Eye, A. (2002). *Configural Frequency Analysis - Methods, Models, and Applications*. Mahwah, NJ: Lawrence Erlbaum.

Wu, H., & Leung, S.-O. (2017). Can Likert scales be treated as interval scales? - A simulation study. *Journal of Social Service Research*, *43*, 527 - 532. doi: 10.1080/01488376.2017.1329775