Journal of Modern Applied Statistical Methods

Volume 6 | Issue 1 Article 3

5-1-2007

Bimodality Revisited

Thomas R. Knapp
University of Rochester; The Ohio State University

Journal of Modern Applied Statistical Methods
May 2007, Vol. 6, No. 1, 8-20

1538 – 9472/07/$95.00

Bimodality Revisited

Thomas R. Knapp
University of Rochester and The Ohio State University

Degree of bimodality is an important feature of a frequency distribution, because it could suggest heterogeneity, such as polarization or two underlying distributions combined into one. The literature contains several measures of bimodality. This article attempts to summarize most of those measures, with their attendant advantages and disadvantages.

Key words: Bimodality, kurtosis, moments, polarization

Introduction

The bimodality of a frequency distribution is of considerable interest in a number of disciplines. A Google search on 'bimodality' returns almost 300,000 entries. Applications of bimodality considerations are found in substantive investigations in fields as diverse as agriculture (e.g., Doehlert, et al., 2004), economics (e.g., Esteban & Ray, 1994), linguistics (e.g., Spivey, Grosjean, & Knoblich, 2005), medicine (e.g., Lim, Bakri, Morad, & Hamid, 2002; Grandi, et al., 2005), psychology (e.g., Lindner, 1997; Beach, Finchman, Amir, & Leonard, 2005), and sociology (e.g., DiMaggio, Evans, & Bryson, 1996; Greeley, 1997; Evans, Bryson, & DiMaggio, 2001; Evans, 2003; Mouw and Sobel, 2001).

Thomas R. Knapp, Ed.D. (Harvard, 1959), is Professor Emeritus of Education and Nursing, University of Rochester and The Ohio State University. He has contributed previous articles to *JMASM* and has a chapter on "Effective sample size" in *Real data analysis* (2007, Information Age Publishers). His website is www.tomswebpage.net; his e-mail address is tknapp5@juno.com.

Esteban and Ray (1984) were concerned with the concept of societal polarization. They argued that one of the indicators of polarization is the bimodality of a frequency distribution for any variable that is an operationalization of an opinion construct such as attitude toward abortion. DiMaggio, Evans, and Bryson (1996), Greeley (1997), Mouw and Sobel (2001) studied the bimodality of several attitude variables--mostly Likert-type scales in the National Election Study (NES) and General Social Survey (GSS) data sets.

Purpose

The purpose of this article is to trace the methodological foundations of bimodality, some of the attempts that have been made to measure it, and some of the contributions to statistical inferences regarding it.

Historical Review

Karl Pearson

In his first of a series of articles on the mathematical theory of evolution, Pearson (1894) devised a procedure for determining whether or not a frequency distribution could be resolved into two normal distributions. The procedure involved six equations in six unknowns (the mean, standard deviation, and membership proportion for each of the two underlying normal distributions), which in turn led to a ninth-degree polynomial equation. If the given distribution had two peaks that were rather far apart it could be described as bimodal. He used as an example some data collected by Professor W.F.R. Weldon on 1000 crabs in Naples.

In a later article (1929) he showed that $b_2 - b_1$, where b_2 is the standardized fourth moment around the mean and b_1 is the square of the standardized third moment around the mean, must be greater than or equal to 1, with the equality holding for the two-point Bernoulli distribution, which is the most extreme case of bimodality.

Darlington to DeCarlo

Darlington (1970) claimed that b_2 (he called it k) is more a measure of unimodality vs. bimodality than a measure of peakedness vs. flatness as often discussed in statistics textbooks, i.e., it is a measure of the extent to which a distribution's z-scores cluster around +1 and -1, with the two-point Bernoulli distribution being the most bimodal, having a k of 1.

Chissom (1970) discussed various interpretations of the kurtosis statistic $\alpha_4 = b_2 - 3$, which is equal to 0 for the normal distribution. He pointed out that $\alpha_4 = -2$ for perfectly bimodal distributions.

In a brief note, Hildebrand (1971) expressed general agreement with Darlington, but gave examples of two bimodal distributions, for one of which k-3 was equal to -1.2 and for the other of which k-3 was equal to 3.

Moors (1986) agreed that k should be interpreted as the extent to which scores cluster around one s.d. to the right of the mean and one s.d. to the left of the mean.

Ruppert (1987) provided a long discussion of the various interpretations that have been made of b_2, including peakedness and tail-thickness, and emphasized Hampel's (1974) influence function approach to the understanding of kurtosis.

Balandra and MacGillivray (1988) wrote a critical review of the literature on kurtosis and favored the viewing of kurtosis as "a vague concept" (p. 116) regarding the location of a distribution's shoulders vis-a-vis its center and its tails.

In a more recent review of the literature on kurtosis, DeCarlo (1997) clarified the role of measures of kurtosis in tests for normality, tests for bimodality, and other matters, in the context of several previously-cited examples.

Reschenhofer and Schilling, Watkins, & Watkins

It has often been claimed that a mixture of two normal distributions is necessarily bimodal. Reschendofer (2001) showed that to be true only if the two modes differ by two or more standard deviation units. Schilling, Watkins, and Watkins (2002) made the same claim for the special case of the distribution of adult heights when men and women are included in the same distribution. Those results are consistent with the arguments made by Darlington (1970) and Moors (1986) regarding the clustering of data at z-scores of +1 and -1 (a difference of two σ's).

Choonpradub & McNeil

Choonpradub and McNeil (2005) were concerned that traditional box plots don't provide any indication of bimodality for the distributions such plots are meant to summarize. They recommended an enhancement (thickening the ends of the box denoting the quartiles) that might reflect bimodality.

Haldane to Frankland and Zumbo

The previously-cited authors were concerned primarily with the description of bimodality. Haldane (1952), however, suggested a fairly simple test for statistically significant bimodality, based upon the successive discrepancies of frequencies for adjacent categories in a sample frequency distribution. He used as an example the distribution of differences in hair color for 162 pairs of siblings.

Shenton & Bowman (1977) laid the groundwork for statistical inferences based upon the skewness coefficient $\sqrt{b_1}$, the kurtosis coefficient b_2, their respective univariate sampling distributions, and their joint bivariate sampling distribution.

A truly bimodal distribution should have a reasonably deep dip between the two modes. Hartigan and Hartigan (1985) developed a dip test that could be used to distinguish between unimodality and bimodality.

Tokeshi's (1992) test of the bimodality of a sample frequency distribution is a type of randomization (permutation) test that compares an actual sample distribution with all of the possible ways the total frequency could have been allocated to the various categories that comprise the variable of interest.

The estimation of the number and location of underlying modes for a sample frequency distribution was investigated by Minnotte (1997).

Frankland and Zumbo (2002) provided an SPSS program for distinguishing between a single underlying normal distribution and a bimodal composite of two underlying normal distributions.

Other Methodological Contributions

There is a set of miscellaneous formulas for the CLUSTER procedure in the SAS User's Guide. One of those formulas, derived by Warren Sarle (Personal Communication, 5/10/06), is a formula for the bimodality coefficient:

$$b = [(m_3^2 + 1)/(m_4 + [(3(n\text{-}1)^2)/((n\text{-}2)(n\text{-}3))])]$$

where m_3 is skewness and m_4 is kurtosis. Values of b greater than 0.555 (the value for a uniform population) may indicate bimodal or multimodal marginal distributions. The maximum of 1.0 (obtained for the Bernoulli distribution) is obtained for a population with only two distinct values. Very heavy-tailed distributions have small values of b regardless of the number of modes.

The notation is unconventional, because the m's usually represent the unstandardized moments about the mean (so just substitute b_1 for m_3^2 and b_2 for m_4). Slight variations of it (for large n the term inside the square brackets is often deleted if 3 has not been subtracted from b_2, or replaced by 3 if it has).

There is another statistic that is also called a bimodality coefficient; it is a function of the likelihood ratio for normal distributions vs. mixtures of normal distributions (see Ashman & Bird, 1994 for an application to astronomy).

In his technical article about L-moments, Hosking (1990) claimed that the ratio of two of them "could be interpreted as a measure of tendency to bimodality" (p. 111).

A Personal View of Bimodality

Bimodality should be thought of topologically. If you push down on the peak of a unimodal distribution the frequency curve gets flatter and flatter until it becomes a uniform distribution. If you keep pushing further the curve crawls upward to the left and to the right and ultimately ends up as a two-point distribution. How then to measure the degree of bimodality of an actual distribution? As Pearson (1929), Shenton and Bowman (1977), and others had pointed out, b_2 - b_1 must be greater than or equal to 1, so that b_2 - b_1 should be a reasonable measure of bimodality, because it takes on its smallest value (1), for the two-point Bernoulli distribution, and it takes on its largest value (conceptually infinite) for a distribution with a single tall peak.

That approach was taken in Knapp (1959) and in a subsequent unpublished paper Knapp (1970) in which an attempt was made to derive the sampling distribution of $b_2 – b_1$ for samples from a normal distribution. That attempt was only partially successful because only the first two moments could be derived mathematically (a Monte Carlo approach was used for the rest of the basis for statistical inference), and significant non-normality is not necessarily the same as significant bimodality.

Some Examples of Descriptive Comparisons

Consider the following hypothetical frequency distributions for a variable that ranges from 1 to 11 and for a sample size of 100 (see Figures 1 through 9).

X	f
1	6
2	10
3	11
4	10
5	9
6	8
7	9
8	10
9	11
10	10
11	6

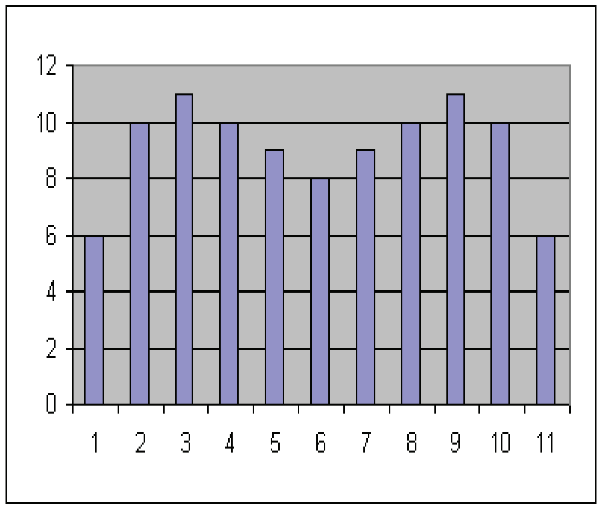

Figure 1

X	f
1	1
2	7
3	9
4	11
5	10
6	8
7	9
8	11
9	12
10	13
11	9

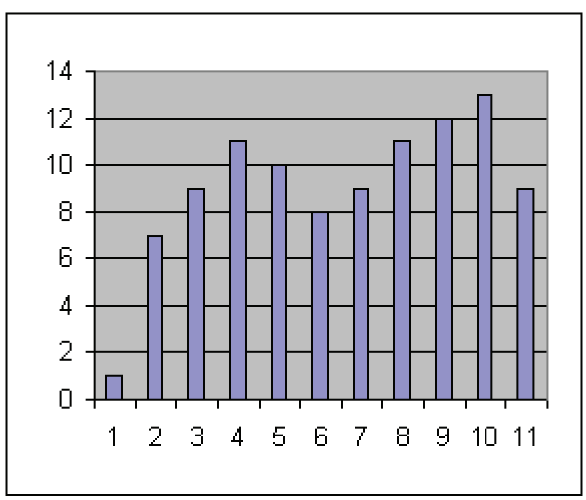

Figure 2

X	f
1	2
2	8
3	10
4	11
5	12
6	14
7	12
8	11
9	10
10	8
11	2

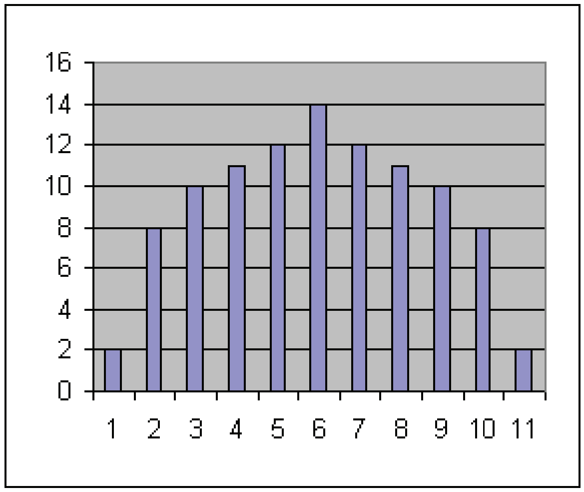

Figure 3

X	f
1	1
2	3
3	6
4	12
5	18
6	20
7	18
8	12
9	6
10	3
11	1

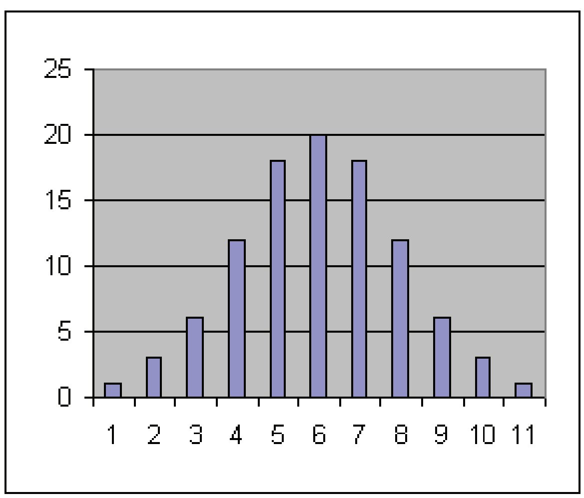

Figure 4

X	f
1	1
2	2
3	4
4	6
5	12
6	16
7	21
8	23
9	14
10	1
11	0

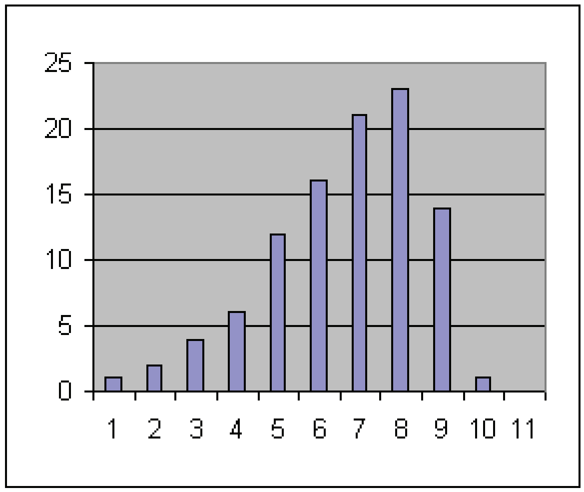

Figure 5

X	f
1	3
2	3
3	4
4	6
5	12
6	44
7	12
8	6
9	4
10	3
11	3

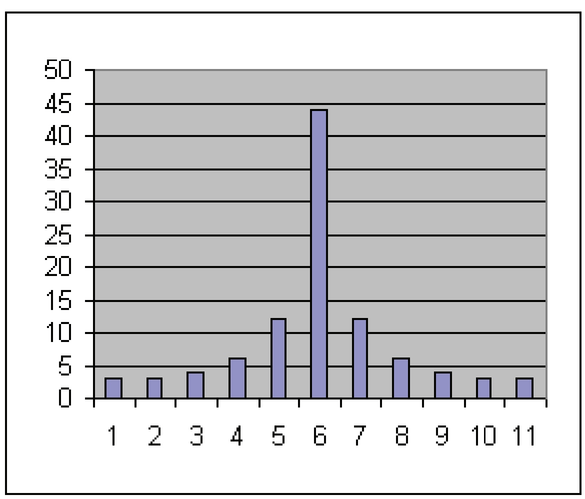

Figure 6

X	f
1	2
2	3
3	3
4	3
5	3
6	4
7	5
8	11
9	21
10	41
11	4

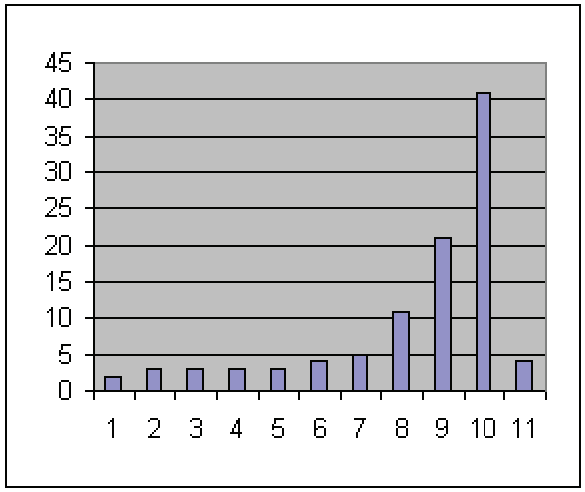

Figure 7

X	f
1	13
2	11
3	10
4	9
5	7
6	1
7	8
8	9
9	9
10	11
11	12

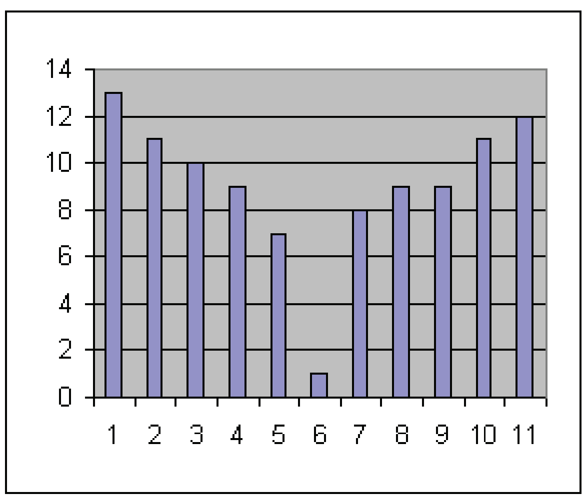

Figure 8

X	f
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10
10	10
11	0

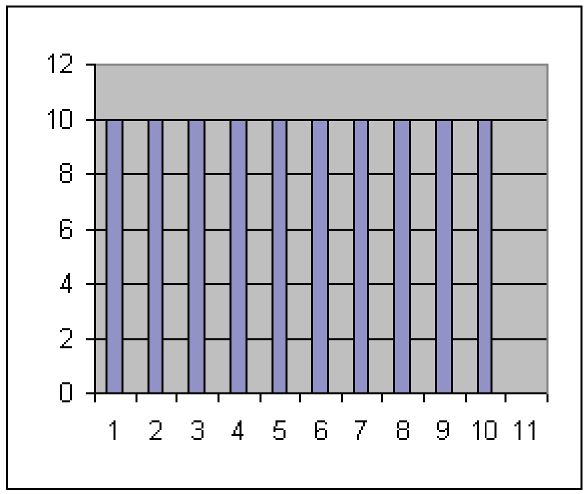

Figure 9

Table 1. Results from Figures 1 through 9

Figure	b_1	b_2	rank*	$b_2 - b_1$	rank*	$(b_1+1)/b_2$	rank*
1	.000	1.757	2	1.757	2	.569	4
2	.021	1.786	4	1.765	3	.572	3
3	.000	2.046	5	2.046	6	.489	7
4	.000	2.804	6	2.804	8	.357	8
5	.549	3.113	7	2.564	7	.498	6
6	.000	4.043	8	4.043	9	.247	9
7	2.443	4.414	9	1.971	5	.780	1
8	.001	1.515	1	1.514	1	.660	2
9	.000	1.776	3	1.776	4	.563	5

Notes. * 1 = most bimodal; 9 = least bimodal

For each distribution, b_2, $b_2 - b_1$, and $(b_1+1)/b_2$ were calculated (see Table 1). The relative agreement among the three measures of bimodality is fairly good except for Figure 7. That figure is clearly not bimodal, which would intuitively rule out $(b_1+1)/b_2$ as an indicator of its bimodality. b_2 alone would suggest that Figure 7 is the least bimodal of the nine figures, but $b_2 - b_1$ would suggest that four of the other distributions (Figures 3, 4, 5, and 6) are less bimodal, with Figure 6 being the least. Therefore, $b_2 - b_1$ is the better indicator of bimodality because a flattening out in Figure 7 may be seen, followed by a second mode popping up at the low end of the scale, if you push down hard enough on the mode at an abscissa value of 10. The appearance of two modes would take much longer with Figure 6 (Three modes would pop up there first--one at each end to go along with the one in the middle).

Statistical Inferences Using the Same Examples

In addition to its simplicity, Haldane's test appears to be the most defensible, because it is appropriate for both interval and ordinal scales. It has been applied to the distributions in Figures 1-9, with the relatively surprising result that none of those distributions is significantly bimodal at the .05 level (see Figure 10). It is Surprising, because Figure 8, for example, really looks bimodal and the sample size is reasonably large (100). But, Figure 10 is an example of one that is; note the deeper trough between the two peaks.

Two real-data examples

Sullivan (2005) found that the frequency distribution of Type 1 rates for age at first birth (with number of previously childless women of childbearing age in the denominator) exhibited a bimodal pattern in the 90s, with peaks at both 20 and 30 years of age. Figure 11 is the Sullivan graph which illustrates that phenomenon for the years 1991, 1995, and 1999:

X	f
1	5
2	5
3	30
4	5
5	5
6	0
7	5
8	5
9	30
10	5
11	5

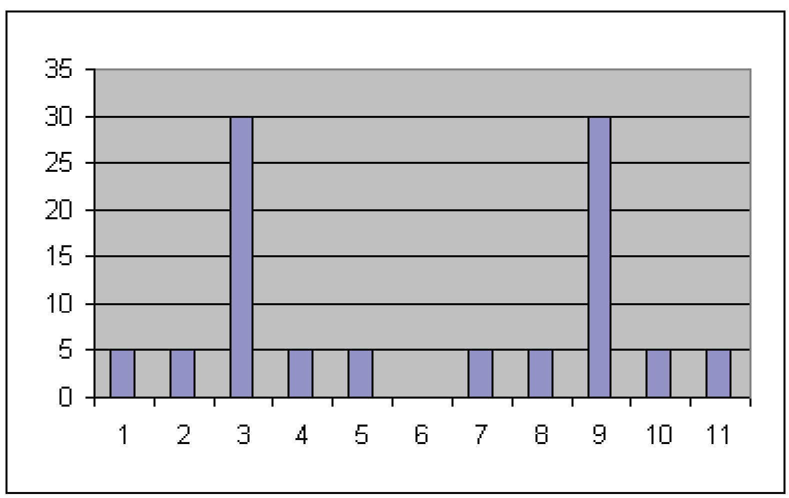

Figure 10

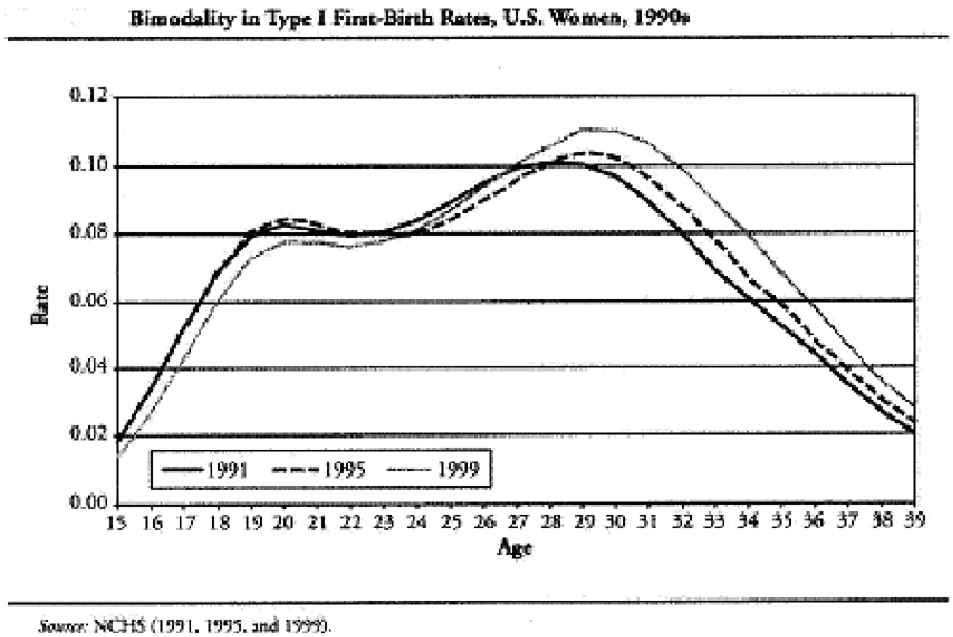

Figure 11. Sullivan Graph

Another interesting recent example of bimodality is discussed in the paper by Roller (2005) regarding the results of a questionnaire sent to U.S. members of the International Reading Association that elicited responses to questions about President George W. Bush's "No Child Left Behind" (NCLB) program. In that article, she said that several of the five-point frequency distributions were bimodal. Here is the example that she emphasized:

Item: "The educational benefits resulting from NCLB implementation in your school district will, on balance, outweigh any adverse impacts for students in the aggregate."

Table 2

Response	Frequency
Strongly Agree	115
Agree	396
Neither Agree Nor Disagree	285
Disagree	357
Strongly Disagree	219
Total	1372
No Response	178
Grand Total	1550

Roller (2005) called the attention of the reader to the modes at Agree and Disagree (see Table 2). b_2 for these data is 1.899; b_1 is .000 (to three decimal places--the distribution is very close to symmetric); $b_2 - b_1 = 1.899$; and $(b_1+1)/b_2 = .527$. Haldane's test supports the hypothesis of underlying bimodality. But the dip between those two modes at Neither Agree nor Disagree could be an artifact of a non-committal response rather than a valley between two peaks. (The large No Response percentage might be further evidence of such an artifact.)

There has been a considerable amount of empirical research regarding the middle category of a five-point Likert-type scale; see, for example, Guy & Norvell (1977) and Armstrong (1987). Mouw and Sobel (2001) argued that DiMaggio et al. (1996) should not have applied their measure of bimodality (b_2 - 3) to Likert-type scales, because it assumes interval-scale properties. The treating of ordinal scales as interval scales is one of the most controversial matters in statistical methodology. There appears to be no solution to the problem that would be acceptable to the warring factions.

Miscellany

Although all of the standard computer packages (SAS, SPSS, Minitab, Excel) include the calculation of one or more measures of skewness and kurtosis, the formulas used in those packages vary somewhat from one another. If you'd like to compute b_1 in Excel, for instance, you need to square SKEW and multiply that by $\{(n-1)^2/(n-2)^2\}$ in order to undo the sample adjustments. As well, in order to compute b_2 you need to add,

$$\{3(n-1)^2/(n-2)(n-3)\}$$

to KURT and multiply that by

$$\{(n-2)(n-3)/(n+1)(n-1)\}.$$

Baretto, Borges, & Guo (2003) pointed out that a typographical error in an article citing one of Tokeshi's (1992) formulas has led to several incorrect tests of the bimodality of distributions that are of interest to researchers concerned with the range-size of various animal species. (Even in their correct form his formulas are tricky, because they require very careful attention to summation operations and combinatorial notation.)

In an interesting article many years ago, Baker (1930) hinted that one should not get too excited about bimodality because a bimodal distribution can often be changed into a unimodal distribution by means of an algebraic transformation. He gave as an example a continuous bimodal fourth-degree polynomial distribution of X that could be converted into a continuous unimodal distribution by replacing X with e^X.

Conclusion

There are several measures of the bimodality of a frequency distribution. There are also several tests of the statistical significance of sample bimodality. Hopefully, this article has provided at least a partial summary of such procedures.

References

Armstrong, R. I. (1987). The midpoint on a five-point Likert-type scale. *Perceptual and Motor Skills, 64*, 359-362.

Ashman, K. M., Bird, C. M., & Zepf, S. E. (1994). Detecting bimodality in astronomical data sets. *Astronomical Journal, 108* (6), 2348-2361.

Baker, G. A. (1930). Transformations of bimodal distributions. *Annals of Mathematical Statistics, 1* (4), 334-344.

Balandra, K. P., & MacGillivray, H. L. (1988). Kurtosis: A critical review. *The American Statistician, 42* (2), 111-119.

Barreto, S., Borges, P. A. V., & Guo, Q. (2003). A typing error in Tokeshi's test of bimodality. *Global Ecology and Biogeography, 12*, 173-174.

Beach, S. R. H., Fincham, S. D., Amir, N., & Leonard, K. E. (2005). The taxometrics of marriage: Is marital discord categorical? *Journal of Family Psychology, 19*(2), 276-285.

Chissom, B. S. (1970). Interpretation of the kurtosis statistic. *The American Statistician, 24*(4), 19-22.

Choonpradub, C., & McNeil, D. (2005). Can the box plot be improved? *Songklanakarin Journal of Science and Technology, 27*(3), 649-657.

Darlington, R. B. (1970). Is kurtosis really "peakedness"? *The American Statistician, 24*(2), 19-22.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2* (3), 292-307.

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have Americans' social attitudes become more polarized? *The American Journal of Sociology, 102* (3), 690-755.

Doehlert, D. C., McMullen, M. S., Jannink, J-L, Panigrahi, S., Gu, H., & Riveland, N. R. (2004). Evaluation of oat kernel size uniformity. *Crop Science, 44*, 1178-1186.

Esteban, J. M, & Ray, D. (1994). On the measurement of polarization. *Econometrica, 62*(4), 819-851.

Evans, J. H. (2003). Have Americans' attitudes become more polarized? An Update. *Social Science Quarterly, 84*(1), 71-90.

Evans, J. H., Bryson, B., & DiMaggio, P. (2001). Opinion polarization: Important contributions, necessary limitations. *The American Journal of Sociology, 106*(4), 944-960.

Frankland, B. W., & Zumbo, B. D. (2002). Quantifying bimodality, Part I: An easily implemented method using SPSS. *Journal of Modern Applied Statistical Methods, 1*(1), 157-166.

Grandi, C., Tapia, J. L., Marshall, G., and the NEOCOSUR Collaborative Group. (2005). An assessment of the severity, proportionality and risk of mortality of very low birth weight infants with fetal growth restriction: A multicenter South American analysis. *Journal de Pedatria, 81*(3), 198-204.

Greeley, A. M. (1997). Polarizing Catholics?: Don't believe your mail! *America, 176*(6), 11-15.

Guy, R. F., & Norvell, M. (1977). The neutral point on a Likert scale. *The Journal of Psychology, 95*, 199-204.

Haldane, J. B. S. (1952). Simple tests for bimodality and bitangentiality. *Annals of Eugenics, 16*, 359-364.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*, 383-393.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics, 13*(1), 70-84.

Hildebrand, D. K. (1971). Kurtosis measures bimodality? *The American Statistician, 25*(1), 42-43.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological), 52*(1), 105-124.

Knapp, T. R. (1959). Two-group classification in the absence of a criterion. Unpublished Ed.D. dissertation, Harvard University.

Knapp, T.R. (1970). A measure of bimodality and its sampling distribution for samples from a normal population. Unpublished paper, University of Rochester.

Lim, T-O, Bakri, R., Morad, Z., & Hamid, M. A. (2002). Bimodality in blood glucose distribution. *Diabetes Care, 25*, 2212-2217.

Lindner, M. D. (1997). Reliability, distribution, and validity of age-related cognitive deficits in the Morris Water Maze. *Neurobiology of Learning and Memory, 68*, 203-220.

Minnotte, M. C. (1997). Nonparametric testing of the existence of modes. *The Annals of Statistics, 25* (4), 1646-1660.

Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician, 40*(4), 283-284.

Mouw, T., & Sobel, M. E. (2001). Culture wars and opinion polarization: The case of abortion. *The American Journal of Sociology, 106*(4), 913-943.

Pearson, K. (1894). Contributions to the mathematical theory of evolution: On the dissection of asymmetrical frequency-curves. *Philosophical Transactions of the Royal Society, Series A, Part 1*, 185, 71-90.

Pearson, K. (1929). Editorial note. *Biometrika, 21*, 370-375.

Reschenhofer, E. (2001). The bimodality principle. *Journal of Statistics Education, 9* (1). [Online]

Roller, C. M. (2005, February) *No child left behind: A survey of its impact on IRA members.* 8 pp. Retrieved from the International Reading Association website, www.reading.org.

Ruppert, D. (1987). What is kurtosis?: An influence function approach. *The American Statistician, 41*(1), 1-5.

Schilling, M. F., Watkins, M. E., & Watkins, W. (2002). Is human height bimodal? *The American Statistician, 56*(3), 223-229.

Shenton, L. R., & Bowman, K. O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and b_2. *Journal of the American Statistical Association, 72*(357), 206-211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *PNAS, 102* (29), 10393-10398.

Sullivan, R. (2005). The age pattern of first-birth rates among U.S. women: The bimodal 1990s. *Demography, 42*(2), 259-273.

Tokeshi, M. (1992). Dynamics and distribution in animal communities: Theory and analysis. *Researches in Population Ecology, 34*, 249-273.