In symmetry is the left the same as the right or is the right left out?

Pamela F. Shaw
Department of Statistics, Macquarie University, NSW 2109, Australia

This study was concerned with finding what characteristics of data, such as direction of skewness, degree of skewness and degree of kurtosis, affected students' ability to use histograms and boxplots for detecting non-symmetry in the parent poplation. The study found that while there was no consistent difference between boxplots and histograms in the proportion of students detecting non-symmetry in the parent population, the direction of skewness did have a significant effect, with more students detecting skewness when the data was displayed in a left-skewed orientation than when the same data was displayed in a right-skewed orientation. This result is consistent with research reported in the psychological literature where many, but not all, studies have shown an over emphasis on the left hand field of view for normal subjects. Other findings of the study are given and suggestions for further research made.

Introduction

Over the years there has been much discussion of the importance of visualisation and use of diagrams in mathematics. Presmeg (1991) has looked at the influence of the teacher on students' use of visualisation. Polya (1957) considers that drawing a diagram should be the first step in problem solving yet Vinner (1989) states that in calculus students are reluctant to draw a diagram.

In statistics Tukey (1977) states that "the greatest value of a picture is that it forces us to notice what we never expected to see"; a view confirmed by Wainer (1992) with his illustrations of vulnerable parts of aeroplanes and sources of typhoid infections in Britain.

Until recently, with the exception of work done by Vernon (1946,1951) and subsequently criticised by MacDonald-Ross (1977), very little work has been done on how students use and understand the graphical displays used in statistics. Recently, with Chance and Data forming a stream in the mathematics curriculum in many countries, this situation is changing.

Bright and Friel (1996) have looked at how children in grades 6 and 8 interpret stem and leaf plots and make connections with histograms and in separate paper have tried to construct a theory of graphicacy.

Before embarking on any formal analysis of data that they have collected students in statistics are encouraged to display their data. That is, they are recommended to create some sort of visual representation of their data. This enables the detection of unexpected relationships, unusual structure in the data and outliers - an outlier being a value unusually far from the main batch of data. In statistics, diagrams are not an alternative way of looking at data (as might be the situation in calculus) but an intergral part of the process. Data displays, for instance, allow for the possible detection of non-symmetry in the parent population. Symmetry, or more specifically normality, is a common assumption for many statistical tests. However, very little work has been done on what students actually see.

The present study, based on statistical displays - histograms and boxplots - is aimed at determining what factors, such as direction of skewness, degree of skewness and degree of kurtosis, influenced students' decisions as to whether or not a sample could have come from a symmetric population. It also aimed to determine which of boxplots or histograms was better for showing genuine departures from symmetry in the population. Boxplots use a rectangular box to identify the middle fifty percent of values in the sample and also highlight the median and extreme values. Histograms give more detail about the internal variation of the values.

Method

One hundred and twenty-five undergraduate students enrolled at an Australian University participated in the study. There were 54 first-year, 34 second-year and 37 third-year students. All students had successfully completed a basic one-semester course in statistics, with second- and third-year students having completed two or more courses.

Subjects were each presented with 32 graphical displays, each of which was generated using the statistical package Minitab. The displays were based on eight simulated data sets each was presented in a left- and a right-skewed orientation (the right-skewed display being the mirror image of the left. The distributions were chosen to ensure the samples covered a range of skewness and kurtosis values. The coefficients of skewness ranged from 0.06 to 1.3 in absolute value; the kurtosis coefficients (m_4/m_2^2) ranged from 2.4 to 12.9. The 16 right-skewed displays, in increasing order of skewness as measured by the associated p-value are shown in Figure 1.

The displays were randomly arranged on the questionnaire with four figures per page. Four versions of the test paper were used; the versions differed in the order in which the questions were presented. All participants were provided with written instructions which included illustrative examples of samples drawn from symmetric and asymmetric populations. The importance of being able to determine symmetry or otherwise was spelt out and this was highlighted by examples which portrayed the existence of sampling variability. Students were required to decide whether each sample could have come from a symmetric, left skewed or right skewed population. There was no time limit, but most students took between 10 and 20 minutes to complete the test.

Analysis

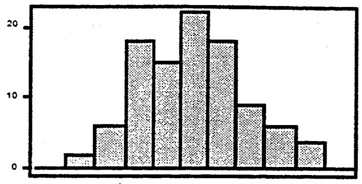
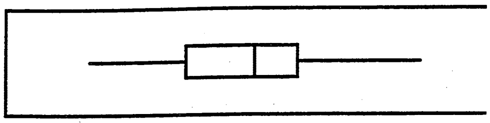
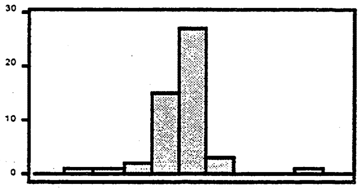
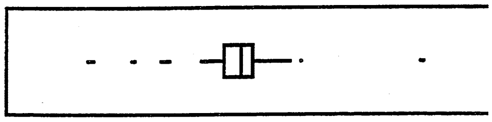
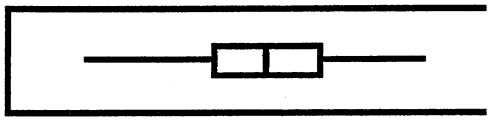
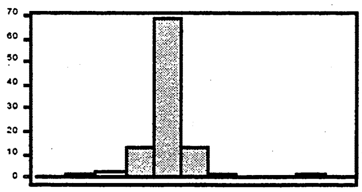
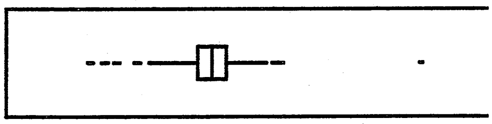
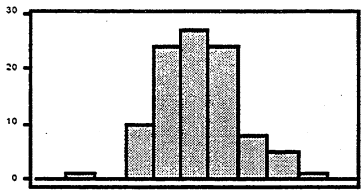
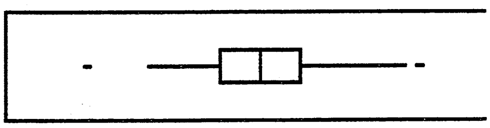
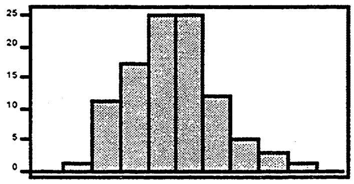
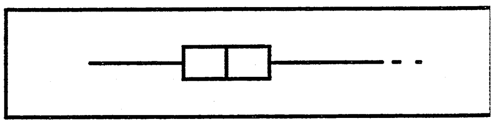
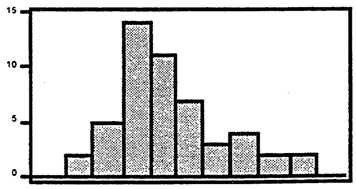
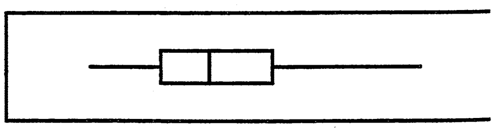
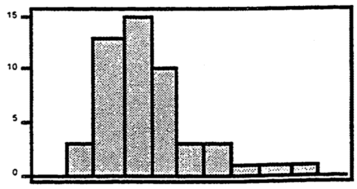
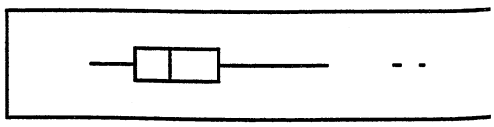
Skewness measures and p-values

The validity of subjects' assessment of skewness can be evaluated by correlating observed outcomes with expected outcomes based on some objective measure of skewness. The skewness of the data sample, not the population from which it was sampled, was the measure used. The sample measures define a characteristic of the display (boxplot or histogram) available to the subject.

The measure of symmetry used was David and Johnson's (1956) statistic

$J = \frac{\hat{\eta} - M}{\hat{\phi}}$ where $\hat{\eta}$ is the average of two symmetric percentiles (the sample minimum and maximum were used), $\hat{\phi}$ is the interquartile range, and M is the sample median. A p-value, or probability of obtaining a larger value of the test statistic if the population was

Figure 1. Right-skewed displays

Distribution	P-value	Histogram	Boxplot
a: n = 100	0.69		
b: n = 50	0.47		
c: n = 50	0.46		
d: n = 100	0.34		
e: n = 100	0.31		
f: n = 100	0.15		
g: n = 50	0.09		
h: n = 50	0.07		

symmetric, was obtained using a bootstrap procedure (Efron, 1979). The bootstrap procedure is based on simulations using the original data. Small p-values are indicative of non-symmetrical populations.

Analysis

On inspection of the data it was apparent that some students were confused at to whether the naming of the direction of skewness was determined by the direction of the tail or the position of the 'hump'. Therefore it was preferable to consider whether or not they considered it symmetrical. For each of the 32 displays the percentage of subjects stating that the sample could have come from a symmetrical population was obtained. These percentages, together with their associated p-values, are displayed in Table 2. These percentages were plotted against the calculated p-values and showed a weak relationship.

Table 2. Percentages of responses : Skewed

Observed (Fitted) Values

Distribution	Sample	P-value (J)	Histogram	Boxplot
a:	100	0.69		
right-skewed			2	39
left-skewed			6	34
b:	50	0.47		
right-skewed			25	28
left-skewed			29	50
c:	50	0.46		
right-skewed			18	18
left-skewed			52	10
d:	100	0.34		
right-skewed			20	57
left-skewed			43	66
e:	100	0.31		
right-skewed			10	11
left-skewed			8	17
f:	100	0.15		
right-skewed			48	36
left-skewed			59	63
g:	50	0.09		
right-skewed			91	90
left-skewed			98	94
h:	50	0.07		
right-skewed			99	98
left-skewed			100	98

A three way analysis of variance of a transformation of the proportion saying the data were skewed was performed. The factors were the direction of skewness, distribution and display type (histogram or boxplot). The transformation was chosen to achieve normality of the response variable.

The analysis of variance found that direction of skewness ($F_{1,15}$=9.52, p=0.008), distribution ($F_{7,15}$=105.80, p=0.000) and interaction between distribution and display ($F_{1,15}$=10.12, p=0.000) were all significant. Subjects showed a greater tendency to say that the displays indicated skewness when the data were skewed to the left than when the same data were displayed in the right-skewed orientation. This was particulary pronounced for distributions that were not highly skewed. Display type was not significant but the interaction with distribution indicated that for some distributions the boxplot gave a greater indication of skewness whereas for others the histogram gave the greater indication of skewness. Further detail of the statistical aspects of this study can be found in Heller et al (1996).

Discussion

This emphasis on the left corresponds to results obtained in psychology. Patients with a lesion on the right hemisphere of the brain, when asked to find the middle of a horizontal line which was placed centrally in front of them, marked the centre to the right of the midpoint. The condition was known as 'neglect' as they were neglecting the left hand part of the display. However, normal subjects when asked to perform the same task marked the middle towards the left. Even though this was not a clinical condition, by analogy with the previous work, it was known as 'pseudoneglect'(Bowers & Heilman, 1980). Possible causes of pseudoneglect have been investigated in relation to line bisection such as hemisphericity of style, (Roig & Cicero,1994), direction of reading (Chokron & De Agostini,1995; Chokron & Imbert,1993), left-right-handedness (Sampaio & Chokron, 1992), hand used (Dellatolas et al, 1996), scanning direction (Brodie & Pettigrew,1996) and sex of subject (Laeng, Buchtel & Butter, 1996; Roig & Cicero, 1994). A similar favouring of the left side was found in a cancellation task using spatial stimuli but was not found with verbal stimuli (Vingiano, 1991). However, Luh (1995) did not find a similar effect in a number of asymmetry tasks. The differences found by Chokron et al would indicate that a repetition of this study with only right handed subjects in a country where initial reading direction was from right to left would be worthwhile.

If the results are indeed due to an overemphasis on the left-hand part of the display this would suggest that even though students have trouble naming the direction of skewness they are influenced in their conclusions by the direction of the tail. Even when the p-value for rejection of the null hypothesis of symmetry was about 0.50 (figures b and c), approximately one-quarter of the students thought the sample had come from a skewed population. It therefore behoves teachers of statistics to give students more examples from symmetric distributions to illustrate the variability that can occur.

The study gives a cautionary warning that until we investigate we can not be certain what students are seeing nor how they are interpreting it

.

References

Bowers,D. & Heilman,K.M. (1980) Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia, 18, 491-498.*

Bright, G.W. and Friel, S.N. (1996)Connecting stem plots to histograms. Paper presented at the Annual.Meeting of the North American.Chapter of the International Group.for the Psychology of Mathematics.Education.

Bradshaw,J.L., Bradshaw,J.A., Nathan,G., Nettleton,N.C. & Wilson,L.F. (1986) Leftwards error in bisecting the gap between two points. Stimulus quality and hand effect. *Neuropsychologia, 24, 849-855.*

Brodie, E.E., Pettigrew.L.E.L. (1996) Is left always right - Directional deviations in visual line bisection as a function of hand and initial scanning. *Neuropsychologia. 34(5) 467-470.*

Chokron,S. & De Agostini,M. (1995) Reading habits and line bisection - A developmental approach. *Cognitive Brain Research. 3(1) 51-58.*

Chokron,S. & Imbert,M. (1993) Influence of reading habits on line bisection. *Cognitive Brain Research. 1(4) 219-222.*

David, F.N. and Johnson, N.L. (1956). Some Tests of Significance with Ordered Variables, *J.R. Statist. Soc.*, B 18, 1-20.

Delatolas,G., Coutin,T. & De Agostini,M. (1996) Bisection and perception of horizontal lines in normal children. *Cortex, 32(4) 705-715.*

Efron, B. (1979). Bootstrap Methods: Another Look at the Jacknife, *Annals Statistics*, **7**, 1-26.

Friel, S.N. and Bright, G.W. (1995) Graph Knowledge: Understanding how students interpret data using graphs. Proceedings of the 17th annual meeting of the North America Chapter of the International Group for Psychology in Mathematics Education. Owens, D.T., Reed, M.K., and Millsaps, G.M.(Ed) Columbus, Ohio: Ohio State University.

Friel, S.N. and Bright, G.W. (1996) Building a theory of graphicacy: How do students read graphs? *Paper presented at the Annual.Meeting of the AERA.1996, New York.*

Heller,G.Z., Hudson,M., McNeil,D., Middeldorp,J., Petersons,M.V., & Shaw,P.F. (1996) The Perception of Skewness. *Technical Report No. S-96-02 Statistics Dept, Macquarie University Australia.*

Laeng,B., Buchtel,H.A. & Butter,C.M. (1996) Tactile rod bisection: Hemispheric activation and sex differences. *Neuropsychologia,34,11,1115-1121.*

Luh,K.E. (1995) Line bisection and perceptual asymmetries in normal individuals - What you see is not what you get. *Neuropsychology, 9(4) 435-448.*

Macdonald-Ross, M. (1977)How numbers are shown: A review of research on the presentation of quantitative data in texts. *Audio-visual Communication Review* 25(4):359-409.

Polya,G. (1957) *How to solve it: A new aspect of mathematical method (2nd ed).* Doubleday, New York.

Presmeg, N. (1991)Classroom aspects which influence use of visual imagery in high school mathematics.*Proceedings of the Fifteenth PME Conference, (3)191-198.*

Roig,M. & Cicero,F. (1994) Hemisphericity style, sex, and performance on a line bisection task - An exploratory study. *Perceptual & Motor Skills, 78(1) 115-120.*

Sampaio,E. & Chokron,S.(1992) Pseudoneglect and reversed pseudoneglect among left-handers and right-handers, *Neuropsychologia, 30(9) 797-805*

Springer,S. & Deutsch,G. (1989) *Left brain, right brain.* New York, W.H. Freeman.

Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley.

Vernon, M.D.(1946) Learning from graphical material. *The British Journal of Psychology* 36(3):145-158.

Vernon, M.D.(1951) Learning and understanding. *Quarterly Journal of Experimental Psychology* 3:19-23.

Vingiano,W. (1991) Pseudoneglect on a cancellation task. *International Journal of Neuroscience, 58(1-2) 63-67.*

Vinner,S. (1989) The avoidance of visual considerations in mathematics *Focus on the Learning of Mathematics,11,2,149-156*

Wainer, H. Understanding graphs and tables. *Educational Researcher* 21(1):14-23, 1992.