

Data Analysis and Probability

STANDARD

for Grades

Instructional programs from prekindergarten through grade 12 should enable all students to—

6–8

Expectations

In grades 6–8 all students should—

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

- formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population;
- select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.

Select and use appropriate statistical methods to analyze data

- find, use, and interpret measures of center and spread, including mean and interquartile range;
- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.

Develop and evaluate inferences and predictions that are based on data

- use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken;
- make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit;
- use conjectures to formulate new questions and plan new studies to answer them.

Understand and apply basic concepts of probability

- understand and use appropriate terminology to describe complementary and mutually exclusive events;
- use proportionality and a basic understanding of probability to make and test conjectures about the results of experiments and simulations;
- compute probabilities for simple compound events, using such methods as organized lists, tree diagrams, and area models.

Data Analysis and Probability

Prior to the middle grades, students should have had experiences collecting, organizing, and representing sets of data. They should be facile both with representational tools (such as tables, line plots, bar graphs, and line graphs) and with measures of center and spread (such as median, mode, and range). They should have had experience in using some methods of analyzing information and answering questions, typically about a single population.

In grades 6–8, teachers should build on this base of experience to help students answer more-complex questions, such as those concerning relationships among populations or samples and those about relationships between two variables within one population or sample. Toward this end, new representations should be added to the students' repertoire. Box plots, for example, allow students to compare two or more samples, such as the heights of students in two different classes. Scatterplots allow students to study related pairs of characteristics in one sample, such as height versus arm span among students in one class. In addition, students can use and further develop their emerging understanding of proportionality in various aspects of their study of data and statistics.

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

Middle-grades students should formulate questions and design experiments or surveys to collect relevant data so that they can compare characteristics within a population or between populations. For example, a teacher might ask students to examine how various design characteristics of a paper airplane—such as its length or the number of paper clips attached to its nose—affect the distance it travels and its consistency of flight. Students would then plan experiments in which they collect data that would allow them to compare the effects of particular design features. In addition to helping students design their experiments logically, the teacher should help them consider other factors that might affect the data, such as wind or inconsistencies in launching the planes.

Because laboratory experiments involving data collection are part of the middle-grades science curriculum, mathematics teachers may find it useful to collaborate with science teachers so that they are consistent in their design of experiments. Such collaboration could be extended so that students might collect the data for an experiment in science class and analyze it in mathematics class.

In addition to collecting their own data, students should learn to find relevant data in other resources, such as Web sites or print publications. *Consumer Reports*, for example, regularly compares the characteristics of various products, such as the quality of peanut butter; the longevity of rechargeable batteries; or the cost, size, and fuel efficiency of automobiles. When using data from other sources, students need to determine which data are appropriate for their needs, understand how the data were gathered, and consider limitations that could affect interpretation.

Middle-grades students should learn to use absolute- and relative-frequency bar graphs and histograms to represent the data they collect and to decide which form of representation is appropriate for different

Middle-grades students should formulate questions and design experiments or surveys to collect relevant data.

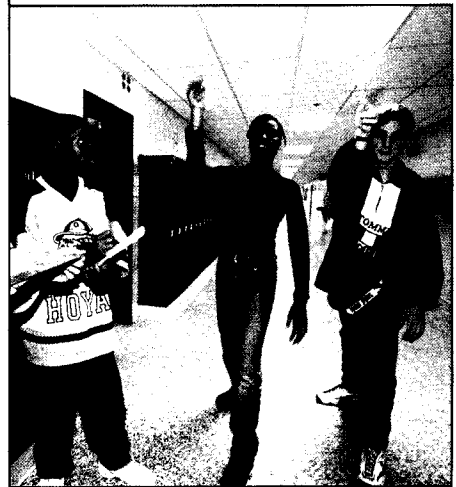
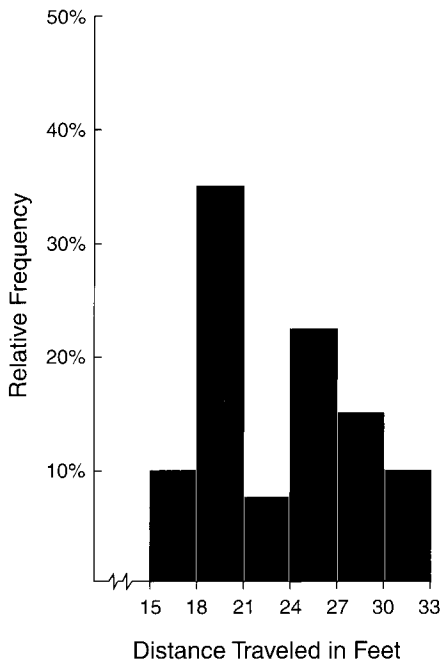




Fig. 6.27.

A relative-frequency histogram for data for a paper airplane with one paper clip



purposes. For example, suppose students were considering the following question:

Compare the distance traveled by a paper airplane constructed using one paper clip with the distance traveled by a plane that is built with two paper clips. Which one travels farther when thrown indoors?

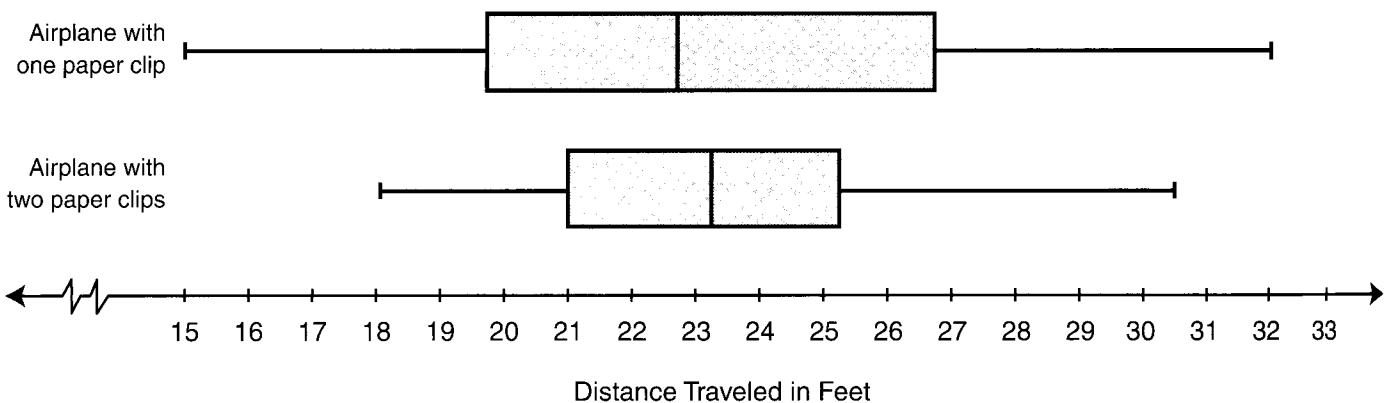
In an experiment conducted to answer this question, one student might throw one of the airplanes forty times while team members measure and record the distance traveled each time. The group might later do the same for the other paper airplane. The teacher might then have the students use a relative-frequency histogram to represent the data, as shown in figure 6.27. For comparison, the teacher might suggest that students display both sets of data using box plots, as in figure 6.28.

Select and use appropriate statistical methods to analyze data

In the middle grades, students should learn to use the mean, and continue to use the median and the mode, to describe the center of a set of data. Although the mean often quickly becomes the method of choice

Fig. 6.28.

Box plots for the paper airplane data



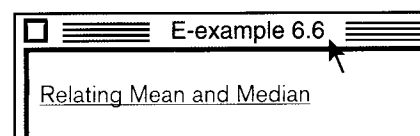
for students when summarizing a data set, their knack for computing the mean does not necessarily correspond to a solid understanding of its meaning or purpose (McClain 1999). Students need to understand that the mean “evens out” or “balances” a set of data and that the median identifies the “middle” of a data set. They should compare the utility of the mean and the median as measures of center for different data sets. As several authors have noted (e.g., Uccellini [1996]; Konold [forthcoming]), students often fail to apprehend many subtle aspects of the mean as a measure of center. Thus, the teacher has an important role in providing experiences that help students construct a solid understanding of the mean and its relation to other measures of center.

Students also need to think about measures of center in relation to the spread of a distribution. In general, the crucial question is, How do changes in data values affect the mean and median of a set of data? To examine this question, teachers could have students use a calculator to create a table of values and compute the mean and median. Then they could change one of the data values in the table and see whether the values of the mean and the median are also changed. These relationships can be effectively demonstrated using software through which students can control a data value and observe how the mean and median are affected. For example, using software that produces line plots for data sets, students could plot a set of data and mark the mean and median on the line. The students could then change one data value and observe how the mean and median change. By repeating this process for various data points, they can notice that changing one data value usually does not affect the median at all, unless the moved value is at the middle of the data set or moves across the middle, but that every change in a value affects the mean. Thus, the mean is more likely to be influenced by extreme values, since it is affected by the actual data values, but the median involves only the relative positions of the values. Other similar problems can be useful in helping students understand the different sensitivities of the mean and median; for example, the mean is very sensitive to the addition or deletion of one or two extreme data points, whereas the median is far less sensitive to such changes.

Students should consider how well different graphs represent important characteristics of data sets. For example, they might notice that it is easier to see symmetry or skewness in a graph than in a table of values. Graphs, however, can lose some of the features of the data, as can be demonstrated by generating a family of histograms for a single set of data, using different bin sizes: the different histograms may convey different pictures of the symmetry, skewness, or variability of the data set. Another example is seen when comparing a histogram and a box plot for the same data, such as those for the one-clip plane in figures 6.27 and 6.28. Box plots do not convey as much specific information about the data set, such as where clusters occur, as histograms do. But box plots can provide effective comparisons between two data sets because they make descriptive characteristics such as median and interquartile range readily apparent.

Develop and evaluate inferences and predictions that are based on data

In collecting and representing data, students should be driven by a desire to answer questions on the basis of the data. In the process, they should make observations, inferences, and conjectures and develop new



The teacher has an important role in providing experiences that help students construct a solid understanding of the mean and its relation to other measures of center.