The learner will understand and use graphs and data analysis.

4

4.01 Collect, organize, analyze, and display data (including box plots and histograms) to solve problems.

Notes and textbook references

A. Price Study Histogram

Have students visit a grocery store to research cookie prices (unit price per ounce), or have them research internet sites or their local video store to get prices of video games. Once the data are collected, have students separate the data into appropriate categories and create a histogram. Keep in mind a histogram has intervals and the bars touch.

At the same time they obtain the data mentioned above, they can record all the different varieties of cookies and video games. They can then survey the class to see how many people like each variety. Using this data they can create a bar graph. Keep in mind a bar graph is used for categorical data and the bars do not touch.

After students have created both a histogram and a bar graph, discuss the difference between the two. How are they different? Similar? What data is used in each one? Can I switch the data and still have them be correct? What other data could I put into these graphs?

B. ESP Graphs - (Blackline Masters IV - 1 and IV - 2)

Materials needed: coin, recording sheet. To start, have students work in pairs. Each student will toss a coin five times. Before each toss, the student will predict the outcome of the toss, heads or tails. At the end of five tosses, the partner records the percentage of times that the prediction was correct. Each partner should do 20 trials of five tosses. After each partner has completed his trials, the students join two other pairs to make a group of six students. The group should combine their data and create a graphical representation that accurately displays their data. Have students present their graphs to the class. Allow classmates to ask questions about the data presented.

Notes and textbook references

C. General Bar Graphs are used to show comparisons of categorical (nominal) data. The vertical axis does not necessarily have to be frequency, but can be any numerical comparison. In a bar graph, the bars do not touch.

Have students choose five or six Asian countries. For each country, find the population, area, literacy rate, and an economic indicator such as GNP. Use each set of data to construct a comparative bar graph for the countries chosen.

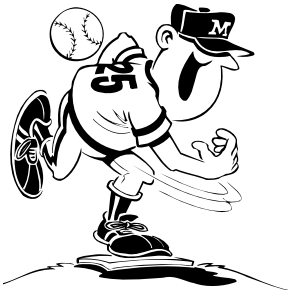
There are various opinions about the best form for a histogram. The following guidelines are often accepted. These guidelines are used to make the graph aesthetically pleasing, easy to read, and they are designed to avoid graphs that can be misleading.

1) A title should be provided for the graph as well as labels for each axis.
2) The bar categories should be listed in numerical order on the horizontal axis. The vertical axis should begin at zero if at all possible. In cases where it is not practical to start the vertical axis at zero, a symbol should be provided to indicate a broken scale.
3) There should be a minimum of about 5 bars in the graph with no more than 20 bars.
4) If each bar represents an interval, the interval widths should be "round" numbers. Depending on the size of the data scale needed, the intervals should count by multiples of 5, 10, 25, 50, 100, 1000 ... some commonly found interval that makes sense for the data scale. Avoid intervals of unexpected width such as 0-7, 8-16, 17-24, etc.
5) The bars should be of equal width.
6) The height of the tallest bar should be somewhat less than (80-90%) the maximum height of the graph grid.
7) If each bar represents a single number, the bar should be centered over the number.
8) If each bar represents an interval of numbers, the left side of the bar should be over the interval minimum, and the right side of the bar should be over the interval maximum.
9) If the data are continous* the bars should touch.

Note: Often histograms are constructed using programs such as Excel or other spreadsheets. These spreadsheet programs may not meet all the guidelines above. Any histogram in which the data are clearly represented, the bar areas are proportional to the number of data points in that category, and the graph is not misleading, is acceptable.
*Continous data are data which can take on any value over an interval. For example, when measuring heights of full-grown tulips, the heights may be anywhere from 8″ to 15″ including fractional heights such as 9.25 inches. No numbers between 8 and 15 are left out. This is an example of continous data.

Notes and textbook references

D. **Batter-Up** (Blackline Master IV - 3 and Blackline Master IV - 13) Students can use baseball statistics to determine the percentage of time that a certain batter hits a single, double, triple, home run, walk or makes an out. They could also determine the percentage of wins and losses a whole team has. Create a circle graph using these percentages. A sample of data is shown on Blackline Master IV - 13. See www.mlb.com for more statistics.



E. **Car Crazy** - (Blackline Master IV - 3)

Have students collect data on the color of cars. They may do this by checking the color of cars in a parking lot, on a car lot, cars that pass a certain intersection, cars they pass on the way home, or merely cars on their street. Have students construct a circle graph of their data. Then combine the data from the entire class and make a circle graph of this combined data. Have each student compare his/her individual chart to the class chart.

F. **Careers** - (Blackline Master IV - 3)

Have all students in the class state whether their careers plan to be in human services, sales, technology, entertainment, research, or manufacturing. Make a circle graph showing the class results. Compare that chart to percentages of new jobs in each category or to current workers in each category.

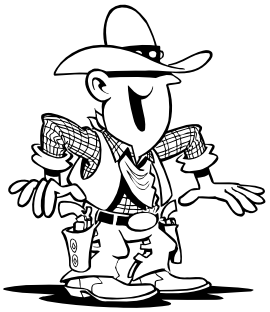
G. **Beanie Babies** (Blackline Master IV - 4)

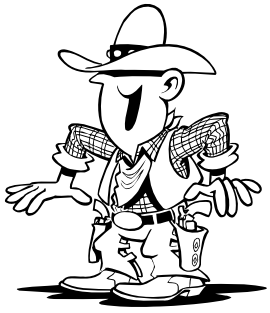
Students do an experiment to determine how many times they have to visit a restaurant to get an entire set of Beanie Babies. This activity is an example of how a simulation can be used to estimate statistical answers. The experimental results can be combined into a box plot to organize and interpret the data. The experiment assumes that the Beanie Babies are packed in the boxes randomly, and that upon purchase any of the animal types is equally likely. Discuss with students how this simulation is alike and different from the real world situation.

Notes and textbook references

H. Comparing Reaction Times (Blackline Masters IV - 5 and IV - 6) Students work in groups to gather data about the reaction time when a meter stick is dropped. Model for students how to position the meter stick and measure the reaction time. Groups should use the data collected to construct a box-and-whiskers plot. Display each group's plot and discuss the plots. Some discussion questions may include:

- Which group had the largest range of reaction times?
- Are there any reaction times that stand out as being very different? How do they affect the box-and-whiskers plot?
- Compare the lower 25% of the data for each group.
- Which group appears to have the fastest reaction times? How can you tell?
- What is the median reaction time for each group? What does it tell you about each group's reaction times?
- Compare the upper 25% of the data for each group.





I. Comparisons of displays of data and which displays are appropriate are important concepts in this objective. Students must understand that certain displays are appropriate for certain types of data. For example – you would not use a histogram to show categories of candy bars. One way for students to understand this is for them to discover it on their own. You can use a multitude of objects: bean mix, candy hearts, fruit loops, M&Ms™, colored blocks to name a few, and have each group sort the contents. They then must construct a chart or graph to display the number of each type. Have each group explain its display. The discussion should lead to the comparisons between the displays and which ones are most appropriate.

Notes and textbook references

J. Commercial Counting (Blackline Masters IV - 14 through IV - 16) Students love to watch TV, so we should capitalize on this fact. This activity can be done over a period of time, or scaled down and done in a class period if you have them collect their data first. (A sheet for data collection is provided.) You can also extend this by asking more intense high level questions.

K. Population Histogram - Have your students research the populations of each African country. Make sure to have student round numbers to millions in order to avoid excessively large numbers. Have them work in groups to create a bar graph of this data.
Ask students the following questions:

1. Which country has the smallest population?
2. Which country has the largest population?
3. What is the population range?

Then have students plan an appropriate histogram to display this data.
Ask students the following questions:

1. What interval width would make sense for this histogram?
2. Which population range is most common for African countries?
3. Which population range is least common?

Have students compare the two graphical representations. What questions are more easily answered using the bar graph? What questions are more easily answered using the histogram?

Many people confuse bar graphs and histograms. While their appearance is similar, there are characteristics specific to each one. Histograms must have numerical data on both the x and y axes. This data is partitioned into consecutive intervals. The bars of a histogram must touch one another since the intervals are consecutive. Bar graphs, on the other hand, can include non-numerical data, like favorite colors, and the bars need not be touching.

L. Mini Review – Graphs (Blackline Masters IV - 24 through IV - 27) This mini review covers most of the skills related to graphing from this unit. Allow students to work in pairs to share strategies and skills.

Notes and textbook references

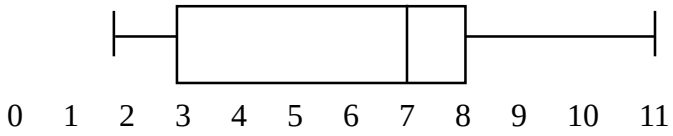
M. **Boxplots** (also called box and whisker plots) are graphical representations of the spread of a data set. Range, the spread from the minimum value to the maximum value in a data set, is shown on a boxplot by looking from end to end. The box in the boxplot represents the interquartile range, the spread from the lower quartile (Q_1) to the upper quartile (Q_3).

Minimum value	lower quartile (Q_1)	median (Q_2)	upper quartile (Q_3)	maximum value

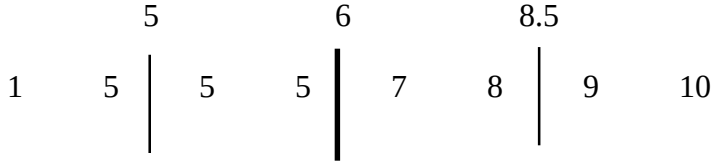
The five number summary used to create a boxplot (minimum, lower quartile, median, upper quartile, and maximum) are determined by splitting the data into four equal groups called quartiles.

For example:

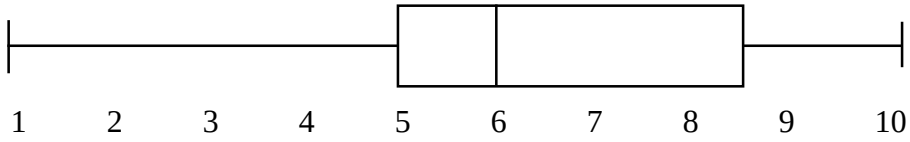
The median is the middle of the data. In this set, the median is 7 because there are three numbers on each side of 7. To find lower quartile, find the "median" of the lower half of the data. In this case, the lower quartile is 3 because there is one value on each side of 3. Similarly, the upper quartile is 8 because 8 splits the top half of the data evenly. The boxplot of this data set would look like this:



If the set has an even number of values, then the five number summary is determined as follows:



The median is the middle of the data. In this set there is no middle number. Thus, the median is the mean of the two middle values, namely 6. This divides the data into two equal pieces because there are four numbers on each side of the median. To find the lower quartile, find the "median" of the lower half of the data. Because there is no middle value, the lower quartile is the mean of the two values in the middle. In this case it is 5. Similarly, the upper quartile is 8.5 because 8.5 splits the top half of the data evenly. The boxplot of this data set would look like this:



Notes and textbook references

Notes and textbook references

4.02 Calculate, use, and interpret the mean, median, mode, range, frequency distribution, and inter-quartile range for a set of data.

A. **Mean, Median, and Mode Bingo** (Blackline Masters IV - 7 through IV - 11 and IV - 23)
Materials needed: regular bingo cards, bingo markers, bingo call cards printed on transparency film and cut apart for use on the overhead, a paper bag or other suitable container for the call cards.

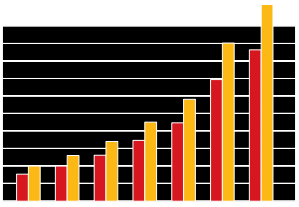
Teachers may find it useful to put an opaque sticky dot on each card and then write the answer on the sticky dot. Using the sticky dot makes the answer easily visible to the teacher, but not to the students. There are 75 call cards which have answers from 1 - 75. The cards are printed with the answers in numerical order.

These bingo call cards each show a set of data and require the student to determine the mean, median, mode, maximum, minimum, or range of the data. Display a card on the overhead; allow time for students to find and mark the answer if it appears on their card. Single line bingo is usually appropriate, but other versions may be played as well.

It is a good idea to allow students to play in pairs so they can discuss the problems. Another tip is to leave one card up as you place a new card on the screen. This allows students who are a bit slower to evaluate the data sets to keep up with the class.

B. Have students get data from home. They might determine

number of garbage cans _____
number of lamps _____
number of telephones _____
number of clocks _____
number of beds _____



In groups of four or five, have students compute the totals and the average for each item in their group. Decide how to record remainders. What do the remainders mean in these situations? Solutions can also be rounded to nearest whole numbers. Then use group totals to determine the class average.
Extension: Determine the average of the group averages. Compare this number to the class average determined using the group totals.

Notes and textbook references

C. **Central Tendency Discussion Cards** (Blackline Master IV - 12) Have students work in groups to discuss the situation described in one or more of the discussion cards. The cards have students evaluate mean, median, and mode and the usefulness of each one as a measure of central tendency in the situation described.

D. **The Average Name** Using small slips of paper or index cards have students write their first names, each letter on a separate slip of paper. Engage the students in a discussion about the average number of letters in a name. Try to have students suggest a way to calculate the average name length for their class without using pencils, paper, or calculators. Lead them to collecting all of the letters from each student into a box. This simulates summing the numbers. To simulate dividing have a student go around and give each child one letter from the box. Continue until each child has the same number of letters. The remaining letters lead to a discussion about remainders or fractional pieces.

E. Students should be able to solve these types of problems:

1. We bought 2 pounds of fudge for $2.98 per pound and 3 pounds of jelly beans for $1.80 per pound. What was the average cost per pound for this candy?

2. The average of Ricardo's three test scores was 87. If he gets 100 on the next test, what will his average be?

3. Jolene has been looking at her grades. The range is 51 points. Her lowest score is 29 and the mean is 65. Can you figure what her highest score is?

4. Maria has an average score of 82 on five tests. What score does she need on the next test to raise her average to 84?

F. Measure body parts, including arms, (shoulder to fingertip), length of foot, circumference of head, waist to ankle, hand span (with fingers spread). Graph each set of data on a line plot or stem and leaf plot. Discuss range, median, mean, and mode of data. Do at beginning, middle, and end of the year and discuss differences that occur.

Notes and textbook references

G. Model for students how to quickly collect data using a stem-and-leaf plot. Then help them organize the data (order it) and make summarizing statements. For example, draw the start of two stem-and-leaf plots on the board or overhead. Have students estimate the total number of pages they have read (any books) during the past week. Record their data. If the first 12 students gave the following numbers as their estimates, your quick graph would look like the sample on the left: 18 ,6, 9, 4, 15, 21, 19, 37, 28, 24, 33, 15. The data, when organized, would look like the stem-and-leaf plot on the right.

0	6, 9, 4,
1	8, 5, 5, 9
2	1, 8, 4
3	7, 3
4	
5	

0	4, 6, 9
1	5, 5, 8, 9
2	1, 4, 8
3	3, 7
4	
5	

H. Have students gather sport statistics (i.e., baseball cards). With these statistics, rank data from least to greatest and determine range, median, mean, and mode. Use calculator for computation. Sets of data can be plotted and compared using a variety of graphic organizers which include line graphs to study trends, bar graphs, and stem-and-leaf plots. Students can use the data and graphic organizers to make mathematical arguments.

I. Try a "Range of Riddles". A set of 3 mystery numbers has a range of 20. There is no mode. The mean is 16. The highest of the three numbers is 28. What numbers are in the set? (Answer: 8, 12, 28) Challenge students to work in pairs and create more riddles.

J. Survey the class to determine the number of members in each student's family. Students need to define "family" as a class discussion. Identify the range, median and mode. Why might the mean not be a good measure to use? Graph data.

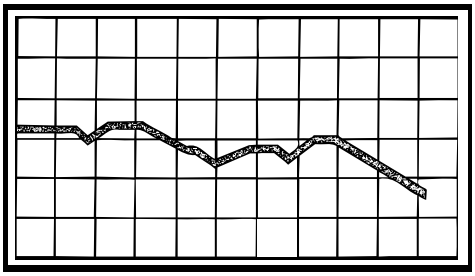
Notes and textbook references

K. Research traffic patterns by the hour either at a local airport (air traffic, arrivals, departures) or in front of school. Graph data and use the information to discuss times of days when traffic is heaviest, the average traffic flow, the range between least traffic and most traffic within a time period. What types of vehicles do you see? What make of car is most common? Are there more foreign or domestic vehicles? What about colors?

L. How much soda do kids drink each month? (How many glasses? how many ounces in a glass?) Have students make a chart and record each morning the number of glasses they drank the day before. At the end of the month, have each student determine a total. Use these data to create a stem and leaf plot. What is the range? the median? the mode? the mean? If each glass costs an average of 28 cents, how much money did this class spend on soda during the past month?

Extension: Do kids drink more glasses of milk or more glasses of soda? Design an investigation to find out.

M. Pretend that you are a stock broker. One of your clients would like you to invest her money by selecting a stock from the New York Stock Exchange. Record the progress of the stock that you selected for ten trading days. Graph the progress of your client's stock. Then, write a report to your client informing her of the stock's progress including the range, median, and mode. Was this a good investment? Explain why or why not.



Notes and textbook references

4.03 Describe how the mean, median, mode, range, frequency distribution, and inter-quartile range of a set of data affect its graph.

A. Notes on Measures of Central Tendency:

Mean: The arithmetic average

Advantages: It is familiar to many people, works well with other statistical formulas, uses and gives weight to all the information in the data set.

Disadvantages: It takes time to calculate and is sensitive to outliers.

Situation where mean is appropriate: When each measurement needs to be included in calculations.

Median: It is the middle score when the scores are listed in order. The median is the same as the 50th percentile.

Advantages: Easy to find and not sensitive to outliers

Disadvantages: The median uses only information about rank (order). It does not give full weight to each point of data.

Situation where median is appropriate: When the impact of outliers needs to be minimized.

Mode: The mode is the most frequently used score.

Advantages: Easy to find and it is not at all affected by outliers.

Disadvantages: The mode may vary greatly from sample to sample and may or may not be anywhere near the middle.

Situation in which the mode is appropriate: When the numbers in the data are more to indicate categories than rank or weight, the mode may be appropriate. It is also useful when looking at grouped data. The modal class is often more a measure of the middle than the true mode of individual scores is.

Example: On a test it may be found that more students score 70-80 than in other ten-point ranges.

B. One of the video tapes in the ***Modeling Middle School Mathematics Series*** is called *Behind the Number*s. In this tape, video clips are shown of students helping a coach decide who should get an MVP award based on their statistics. Mean, median, mode, and range are used in the decision-making process. To find out more about this tape, visit the website: http://www.mmmproject.org/ and click on the video matrix link near the top of the page.

Notes and textbook references

C. Play a mental math game with a set of data. Put a set of data up on the board. Have students graph the set of data. Then call out a number that you would add to the data. The students must tell you what will happen to the graph (without graphing it). Then, you can graph it to check their answer. Students can develop a number sense with regard to the mean and median.

Many students may be able to define and calculate the mean, median, and mode of a set of data. However, when it comes to analyzing what those results mean, they sometimes fall short. This objective is meant to allow the student to understand how adding or subtracting data from the original set will change the mean, median, mode, range, frequency distribution, and inter-quartile range; and therefore affect its graph.

D. **Tinker Plots ™-** Tinker Plots™ is software that introduces a new way of making graphs – by using a construction set of graph pieces that students add one at a time to build what they want. Students can explore their graphs to identify measures such as the mean, median, and mode. In addition, they can drag an outlier and watch these measures change. See the website: http://www.umassd.edu/scrmp/websites.cfm for ordering info and other additional info.

E. **How much do you make?** (Blackline Master IV - 17)
This activity will take your students through the process of analyzing their data and graphing it. They will then see what happens when you add a very high number to the data. Discuss these outliers and how common they are in certain data. For example, in salaries and home prices, these outliers are very common and therefore the median is used to describe salaries and home prices, and not the mean. In addition, you could have them go to the internet and get listings of the median home price and salary.

F. Students enjoy working with their own data. You can use data such as: student height, number of siblings, averages from a test, number of TV shows watched this week, ages of people, etc. Pick a type of data, gather the data, and have the students go through the following questions:

1. Make a graph for the set of data (pick appropriate to the data.)
2. Determine the mean and median for the set of data. (And more if you want)
3. Explain which measure best describes the data and why.
4. Have them add some new data in and have them explain what happens to the data – both numerically and graphically.

Notes and textbook references

G. The most dramatic example of a set of data affecting a graph that relates to the student is scoring a zero on an assignment.

Give the students all of their scores they have received for the quarter (or give them all the same set if that is easier) and have them calculate the mean, median, mode, range, frequency distribution, and interquartile range. Then have them graph the data in as many graphs as they can. Have them then add in a zero to the data and recalculate all the numbers and redraw the graphs. Discuss how this outlier can completely affect their grade in a negative way.

Extensions:

1. If they have a zero, how many 100's do they need to get to bring their grade back up to what it was?
2. Will a score of 100 have as much impact on the graph as the zero does?
3. Will a score of 50 (or 40 or 60) have as much impact on the graph as zero does?

4.04 Identify outliers and determine their effect on the mean, median, mode, and range of a set of data.

Notes and textbook references

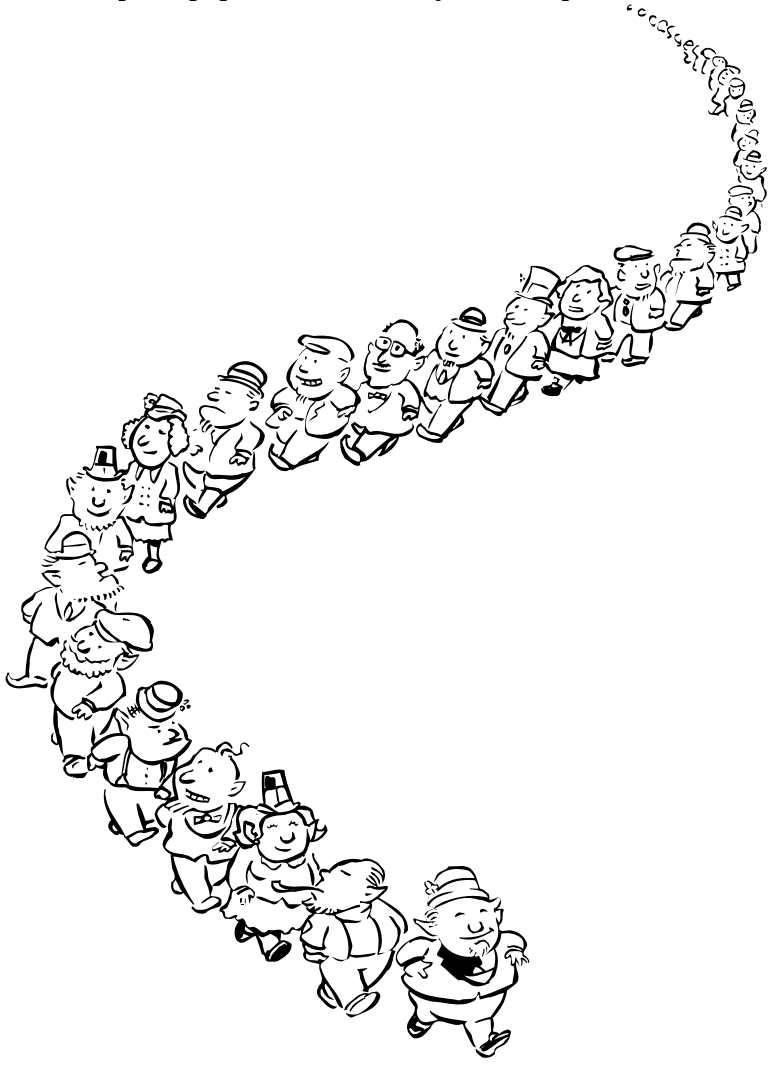
A. A good way to get students thinking about outliers and their effect on measures of central tendency is to discuss grades. This topic is something that students are familiar with and can relate easily to. Give students a list of seven to ten grades and have them calculate the measures of central tendency for the data set. Ask students to consider that they did not hand in a particular assignment and received a grade of zero. Allow students to discuss what they think will happen to the measures of central tendency when this grade of zero is factored into the data set. Have them re-calculate the measures of central tendency this time including the zero. Using other sets of data, lead students into a discussion about how an outlier that is far below the lower quartile will always lower the measures of central tendency when included. Consider and investigate the opposite situation as well.

B. **Using Spread Sheets** The use of spread sheets is a tool that easily allows students to see the effect that an outlier has on a data set. Provide students with a made up set of data or have them collect data. This would be a great opportunity to collaborate with the science teacher on your team. Using a spread sheet application, have students input the data and set up columns for the mean, median, mode and range. Encourage students to input a new piece of data that is unusually larger or smaller than the other pieces of data. Have them observe how the columns for mean, median, and range change. Lead a discussion and have students arrive at generalizations about the effects of outliers that are lower numbers as opposed to outliers that are higher numbers.

C. A great way for students to build mathematical understanding of statistical measures is through the process of collecting and analyzing data. Collaborate with the science teacher on your team for this. When the students are completing a science lab have them use the numerical data they collected to calculate the measures of central tendency. Invariably, a student will make a mistake in the data collection process or not follow the lab procedures correctly and an outlier will be present. Discuss with the students what effect the outlier has on the data. Have the students calculate the measures of central tendency with and without the outliers and analyze the differences.

Notes and textbook references

D. Human Number Line To provide visual and kinesthetic learners with a way to build understanding of outliers in a data set and their effect on the measures of central tendency, try using a human number line. Distribute index cards with one member of a data set written on each card. Have students arrange themselves numerically in the front of the classroom. Have students explain the importance of arranging the numbers numerically and then have them identify the mean, median and mode. Give the student(s) that represents the mean, median and mode(s) a colored index card labeled appropriately. If these values are not part of the data set, allow a student to write the value on the appropriately labeled card and move into the line. Now include a new member of the data set that is clearly an outlier. Ask the students if the measures of central tendency will remain the same or change. Once the students have identified the new measures of central tendency, have a student go up and physically move the appropriate labels. As the student is moving the labels, discuss with the students what is happening to these values as outliers are added into the data set. Repeat this process several times with different outliers. For students who may need to do each one of these steps individually, provide the numbers and labels written on small slips of paper that can easily be manipulated at a desk.



4.05 Solve problems involving two or more sets of data using appropriate statistical measures.

Notes and textbook references

A. **Who's the Best? Activity** (Blackline Masters IV - 18 through IV - 21) This activity asks students to compare data for four basketball players in order to determine who is the best. Students begin by calculating and comparing the basic statistical measures for each player. Then, they use graphical representations to compare these players. The activity concludes with students presenting their findings. Also included in the Blackline Masters is an activity that allows students to use the TI-83 or TI-84 graphing calculators to graph boxplots.

B. **Comparing Test Scores of Two Classes (**Blackline Master IV - 22) In this activity, the students are asked to compare test scores from two different classes displayed in a back-to-back stem and leaf plot and draw conclusions from the data.

C. **Jump Rope** (Blackline Masters IV - 28 and IV - 29) The blacklines provide data collected from two classes regarding the number of consecutive jumps recorded when jumping rope. Without being told what mathematical calculations to make, students are asked to analyze the data to answer questions. Allowing students to complete this activity in small groups provides opportunity for good discussions among the students about what things need to be calculated and how to use those calculations to interpret the data.

D. Students often become more interested in data when it pertains to them. A great way to encourage students to investigate multiple data sets is to have them design experiments that require collection of more than one set of data. Students should devise questions that can be answered through the analysis of the data sets before they perform the experiment. Have students write statements that compare and contrast the data or give presentations with the use of graphical representations that describe the experiment conducted and the questions answered.

Notes and textbook references