Analyzing and Interpreting Scientific Data

How can analyzing and interpreting scientific data allow scientists to make informed decisions?

Why?

During scientific investigations, scientists gather data and present it in the form of charts, tables or graphs. The data must be properly collected, analyzed, and interpreted to allow scientists to make informed decisions regarding the validity of their study and any further work that may be necessary to achieve their objectives. The ability to present and use data charts, tables, and graphs correctly is essential for good scientific practice and also prevents unnecessary or inappropriate work and misinterpretation of the data.

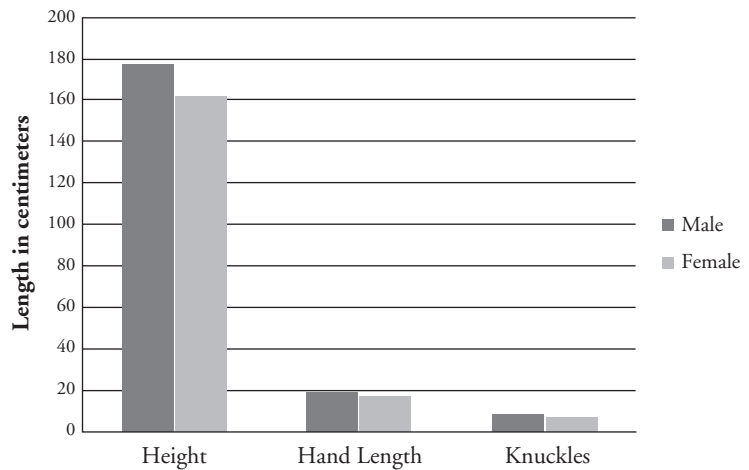## Model 1 – Graphs and Charts of Classroom Measurement Data

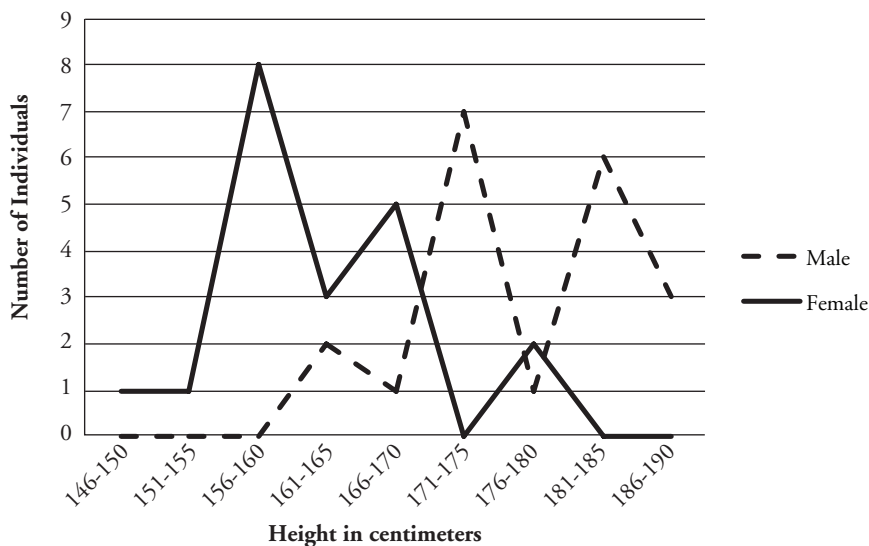### Pie Chart

Percentage of Males and Females by Height
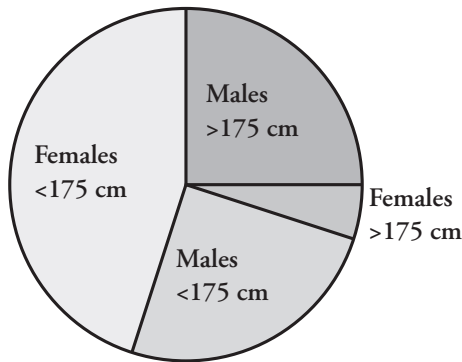
Bar Graph

Comparing Male and Female Average Values

Line Graph

Distribution of Height in Males and Females

1. According to the data in Model 1, how many females fall within the range 146–155 cm tall?

2. According to the data in Model 1, how many males are 181 cm or above in height?

3. Using the graph(s) in Model 1, determine the approximate average height of males and of females.

4. Refer to the data in Model 1.

   a. How many males are taller than 175 cm and approximately what percentage of the total is that?

   b. Which graph(s)/chart(s) illustrate the answer to the previous question?

5. Which type of graph or chart in Model 1 shows a side by side comparison of data?

6. Which type of graph or chart in Model 1 shows trends in data across an entire data set?

7. Describe two trends in male and female height using the line graph.

8. Use complete sentences to compare the presentation of height data in the three graphs. Discuss any information that is located on more than one graph, and any unique information that is available on each.

9. If you wanted to see if a correlation exists between the height of an individual and his/her hand length, what would be the best type of graph/chart to make? Explain your reasoning.

10. What conclusions can you draw comparing the height, hand length, and knuckle width of males and females? State your conclusions in complete sentences.

Model 2 – Foot Width in a High School Classroom

Female foot width (cm)	Male foot width (cm)
7.8	10
8	10.5
8	9
5	9.3
17	13
7.5	7.5
7.5	10
7	9.2
7.8	9
7	4.5

Mean $= \dfrac{\text{sum of all data values}}{\text{number of data values}}$

Median $=$ Middle value of an ordered set of data.

Mode $=$ Most frequently occurring value in a set of data.

11. Refer to the data in Model 2.

    a. What value for foot width is most frequent in males?

    b. What is this value called?

12. Determine the median value for foot width for males and for females. Describe in complete sentences the method you used to determine the median values.

13. Determine the mean for each data group, and describe in a complete sentence how you calculated them.

STOP

Read This!

Within a data set there may be individual values that seem uncharacteristic or do not fit the general trend. These data points may be referred to as **outliers** or **anomalous data**. In most samples, a small number of outliers is to be expected, due to the variation inherent in any naturally-occurring population. Outliers can also result from errors in measurement or in the recording of data. Normal variation can often be distinguished from error by repeating the measurements to see if the same range is obtained. Scientists also use statistical calculations to determine the expected range of data, so that judgments can be made about the authenticity of individual data points. Outliers should not be ignored, however, as many interesting scientific discoveries have resulted from the study of such unexpected findings.

14. Which data point(s) in the foot width values in Model 2 might be considered outliers? Explain your choice(s).

15. The equation below allows you to calculate the amount of deviation (in percent) for the values within a data set. The percent deviation is reported as an absolute value.

$$\% \text{ deviation} = \frac{|(\text{mean value using all data}) - (\text{mean value excluding anomalous data})|}{\text{mean value using all data}} \times 100$$

a. What is the percent deviation in the female data set when the outlying value of 17 is excluded (*i.e.,* considered to be anomalous data)?

$$\% \text{ deviation} = \frac{|8.26 - 7.29|}{8.26} \times 100 = 11.7\%$$

b. What is the percent deviation in the male data set when the outlying value of 4.5 is excluded?

$$\% \text{ deviation} = \frac{|9.20 - 9.72|}{9.20} \times 100 = 5.65\%$$

c. Which data set (male or female) had the largest percent deviation?

16. Given the outliers and amount of deviation in each data set, which value (mean, median, mode) *best represents* the overall data set of foot width in males and females? Explain your answer in a complete sentence.

Extension Questions

17. With your group, discuss the issues below relating to data analysis and scientific ethics, and record your answers in complete sentences.

    a. What could you do to determine whether the outliers in Model 2 are authentic measurements?

    b. Under which circumstances would it be appropriate to remove outlying data points from the analysis and conclusions in a scientific study?

    c. If you were to decide to remove outlying data points from your analysis, what are two ways you could indicate this in your report to ensure you are being honest about your data analysis?