Tree height: Averages, medians, and modes

GLOBE		Associated SDG	Type of Activity
Sphere	**Protocols**		
Biosphere	Biometry - Tree Height	15: Life on Land 13: Climate Action 4: Quality Education	Cognitive

Overview

When scientific measurements are made, to be sure that the data taken are as close as possible to reality, they should be repeated at least 3 times. This number of repetitions helps to reduce the error. The data obtained are then summarized using some statistical tools such as averages, modes or medians. Through this activity, students will calculate averages and find the median and mode of a set of data related to tree height.

Time

50 minutes

Prerequisites

None

School level

Secondary

Purpose

Considering tree height data, students will learn the importance of repetitions, calculate averages, and identify the median and mode of a data set.

Students Outcomes

- Students will recognize the importance of replicates when taking scientific data.
- Students will organize a data set to obtain summary statistical measures.
- Students will calculate summary statistical measures of a data set such as mean, median, and mode.

Background

When field measurements are made to obtain data on some observed characteristic, it is important to have confidence in the measurements taken, as only then can the data be properly contextualized and given meaning. The confidence in the measurements depends on some factors such as the state of the instruments used (therefore, the calibration of the equipment is fundamental), the place where the measurement is made, the conditions in which the data are obtained and the subjectivity (experience) of the observer.

The first step for an adequate measurement is to identify the variable to be evaluated, to select the appropriate instrument, to calibrate it and to follow a set of established steps to take the measurement in such a way that it can be repeated on successive occasions. This process is called measurement protocol.

Confidence in the measurements implies recognizing that each time a measurement is repeated, the values obtained will be very close to each other, being within a range or confidence interval. To make sure that the data we obtain are reliable, it is always recommended to make at least 3 measurements of each variable under study to reduce the error. For example, when measuring the height of trees using the GLOBE Observer App, or a clinometer, it is recommended to measure each tree 3 times and make sure that the difference of these three measurements is less than 1 m in height (this meter would be the confidence interval).

Repeating the measurements generates a large amount of data, which often times is not known how to handle, which is why the data from the repetitions are organized and summarized in some measurements that have greater meaning. Within these measurements are the means (averages), medians and modes, which are part of what is known as descriptive statistics.

Statistics can be defined as the science that makes sense of data using mathematics. In order to properly interpret the data on a given reality, it is important to be clear about some concepts, such as the following (INEI, 2006):

- Population: The set of units or elements defined in space or time in which the researcher is interested.
- Sample: The representative subset of the population.
- Sampling unit: The unit that is selected to construct the sample and from which the data is obtained.
- Error: The difference between the population real value and the sample estimated value.
- Average: A measurement of central tendency obtained by adding the values of the data and dividing the result by the number of them.
- Median: A measurement of central tendency that divides a set of sorted data into approximately two parts, 50% higher and 50% lower. To calculate the median, the data are sorted from smallest to largest. If the number of data is odd, the median is the central value. If the number of data is even, there are two central terms, and the median is the semi sum of these two values.
- Mode: A measurement of central tendency that presents the value of the variable that has the greatest frequency. That is, which is the value that is most repeated.
- Data: Value of a variable that in itself has no meaning if it is not properly analyzed.
- Variable: A characteristic of an object that can be observed, measured and analyzed.

Guiding Research Questions

- What is an average?
- Why is it important to calculate averages?
- Is the mean or average the same as the median and the mode?
- Why do the measurements have to be repeated?
- What calculations can be done with the data to better understand it?

Scientific Concepts

- Mean
- Median
- Mode

Materials and Tools

- GLOBE Observer App – Tree Height
- Paper
- A pencil
- A calculator

What to Do and How to Do It´

- ***Beginning***
  - Teachers go into the field with their students and assess the height of the trees they find. It is important to tell the students to select random trees to be measured and to take three measurements of each tree.
  - If they do not have time to go out in the field, students can use Figure 1, which shows some tree measurements.
- ***Development***
  - The teachers present Figure 1 and ask their students to identify the trees with reliable data to calculate the average height. Each tree has been measured three times and it is important to remember that to have confidence in the data, the data must be within one meter of each other.

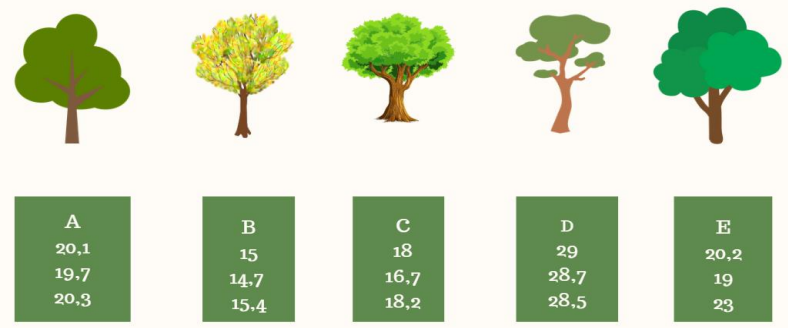

Figure 1: Measurements in meters taken from five trees. It can be seen that trees A, B and D have data with a difference from each other lesser than 1 meter apart, so the data are reliable. In tree C the difference between the first and second data is greater than 1 meter, so one more measurement should be taken. On the other hand, in tree E, the three data among them have differences greater than 1 meter, so at least two new tree E measurements should be taken.

- Once the trees with reliable data have been identified, the students are asked to calculate the average height (Table 1). It should be remembered that, to calculate the average, the three data are added together, and the result of the sum is divided by 3 (because there are 3 measurements).

Table 1
Tree Height Average Calculation

Tree	**A**	**B**	**D**
Datum 1	20,1	15	29
Datum 2	19,7	14,7	28,7
Datum 3	20,3	15,4	28,5
Average			

- Next, the students are asked to look at Figure 2, which shows the average measurement of a series of trees and are asked to calculate the mode and median of the values shown.

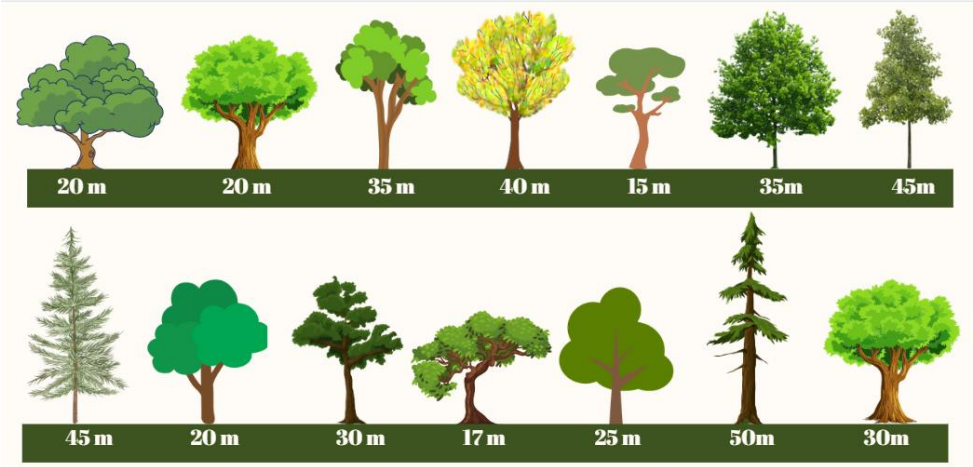

Figure 2: Average of the measurement of a set of trees

- To calculate the mode, it is important to sort the data from smallest to largest and count the number of times that value is repeated- this is called frequency. Therefore, the value of the mode will be the number that appears most frequently, as shown in Table 2. In this example, the mode is the 20 m trees. If there is a tie in the number of times a value appears, it is possible to consider that there are two modes.

Table 2
Mode Calculation

Tree size	15m	17m	20m	25m	30m	35m	40m	45m	50m
Frequency Number of trees with the same size	1	1		1			1		1

- Next, the median of the values is calculated. For this purpose, the tree height sorted data from table 2 is considered, which can be seen in Figure 3. To calculate the median, the middle value of the data set is considered. In this case, if there are 9 tree heights, the median will be the value that is in a central position (middle value). To its left there will be the 4 low values and
to its right there will be the 4 high values.

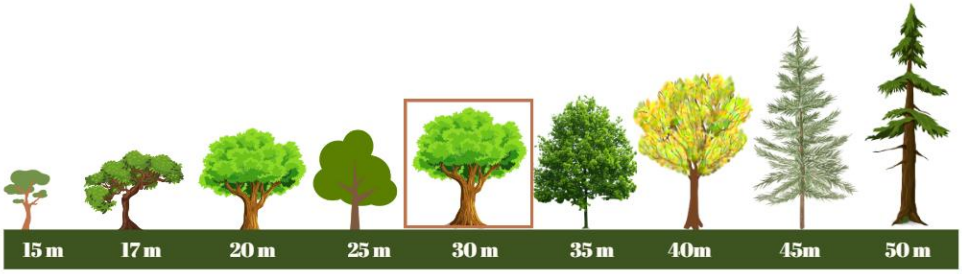

Figure 3: Trees are sorted by height, from shortestt to tallest, with the median shown in a square.

- In Figure 3 the median of an odd set of data has been calculated. What would happen if the data set were even is that there would be two central values so the median would be the average of the two central values.

- ***Closing***
  - The students explain why it is important to repeat a measurement in their own words.
  - Working in groups, the students create a poster to explain what they understand by mean, median and mode, and why these values are important for generating information from data collected in the field.

Frequently Asked Questions

Is it possible to work with other elements in class, if we don't have trees nearby?
Of course, you can use plant leaves that students can bring to class and use a ruler to measure them. You can also have different students measure the same leaf so that they can see for themselves if there is a difference among the measurements.

What exactly is a variable?
The variable is something that can change, adopting different values. These can be qualitative (generally opinions that describe attributes or qualities) or quantitative (numbers that present some dimension of the variable being studied). In turn, quantitative variables can be continuous or discrete. They are called continuous variables when they are described with real numbers showing intervals, for example, height of the trees: 5.53 m or 4.21 m. Discrete variables are those that are represented by natural numbers, such as the number of trees in the yard: 15 or 20 (INEI, 2006).

Variables can also be classified as independent and dependent. Independent variables are those that directly influence the characteristics being studied, for example, time. Dependent variables are those that are measured in each repetition to see how the independent variable influences its values, it can also be said that it is the response variable, for example, the height of the trees, which changes over time (Florian, 2008).

Suggested Resources

- GLOBE Observer Trees Overview - GLOBE Observer - GLOBE.gov
- Students can use Canva to make their posters. Home - Canva

Bibliography

INEI - National Institute of Statistics and Informatics of Peru. 2006. Basic Glossary of Statistical Terms. Retrieved from https://bit.ly/3ANPgwf

Florián, A. 2008. Characterization of the RF Signal Propagation Quarry. Bachelor Thesis. Electronics and Communications Engineering. Department of Computing, Electronics and Mechatronics, School of Engineering and Sciences, Universidad de las Américas Puebla.140 pp.