What is the role of statistics in creating and interpreting box plots?

Authored by

stats writer

June 29, 2024

Statistics plays an essential role in creating and interpreting box plots as it provides a visual representation of a set of data and helps to identify the key characteristics and trends within the data. The process of creating a box plot involves using statistical measures such as the median, quartiles, and interquartile range to summarize the data and draw a box and whisker plot. This allows for a quick and easy understanding of the data distribution, central tendency, and variability. Furthermore, statistics also aid in interpreting box plots by providing insights into the outliers and the overall shape of the data. Overall, the use of statistics in creating and interpreting box plots enables individuals to gain a deeper understanding of the data and make informed decisions based on the insights gathered.

Statistics - Box Plots

A box plot is a graph used to show key features of quantitative data.

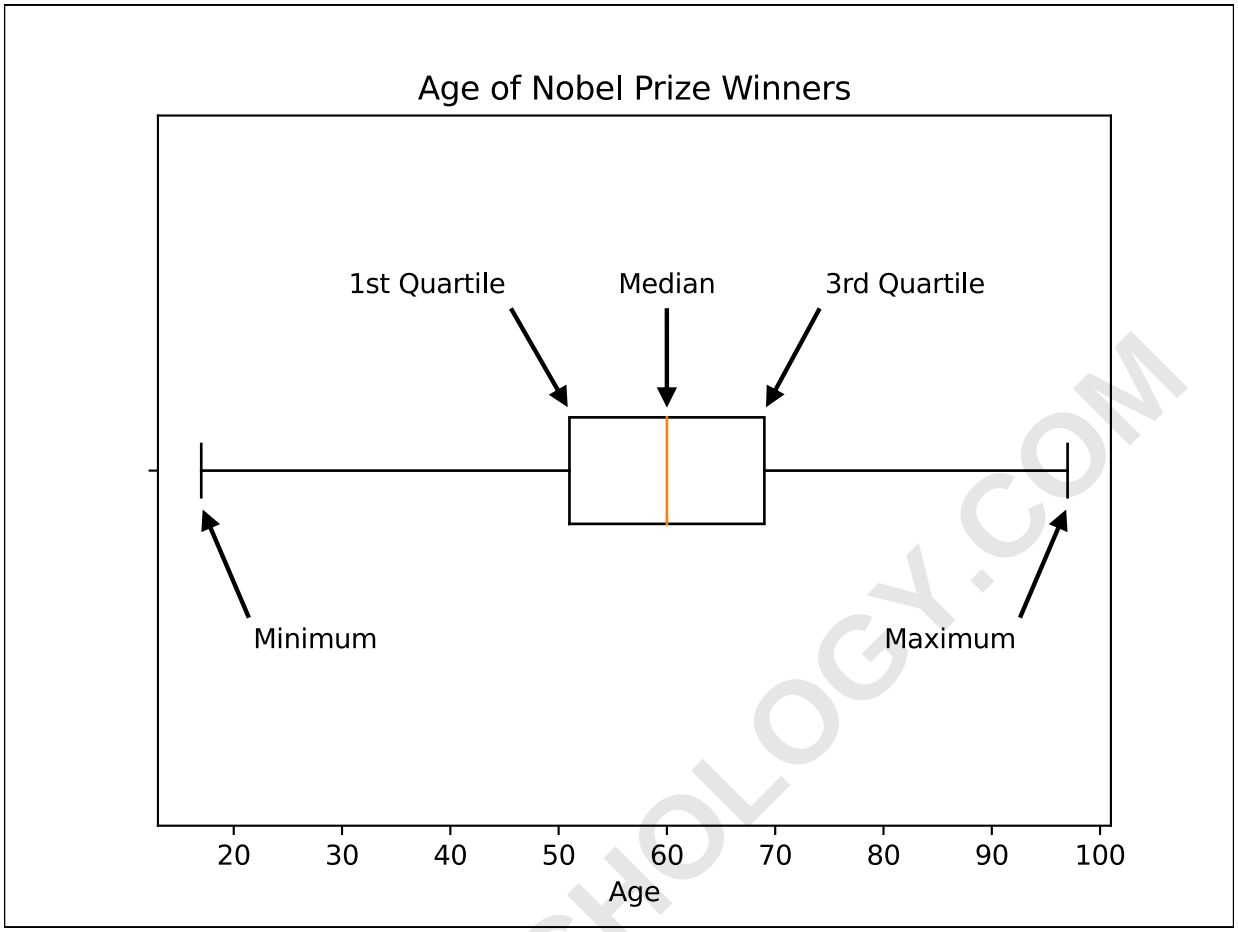
Box Plots

A box plot is a good way to show many important features of quantitative (numerical) data.

It shows the median of the data. This is the middle value of the data and one type of an average value.

It also shows the range and the quartiles of the data. This tells us something about how spread out the data is.

Here is a box plot of the age of all the Nobel Prize winners up to the year 2020:



The **median** is the red line through the middle of the 'box'. We can see that this is just above the number 60 on the number line below. So the middle value of age is 60 years.

The left side of the box is the 1st **quartile**. This is the value that separates the first **quarter**, or 25% of the data, from the rest. Here, this is 51 years.

The right side of the box is the 3rd **quartile**. This is the value that separates the first three **quarters**, or 75% of the data, from the rest. Here, this is 69 years.

The distance between the sides of the box is called the **inter-quartile range (IQR)**. This tells us where the 'middle half' of the values are. Here, half of the winners were between 51 and 69 years.

The ends of the lines from the box at the left and the right are the minimum and maximum values in the data. The distance between these is called the **range**.

The youngest winner was 17 years old, and the oldest was 97 years old. So the range of the age of winners was 80 years.

Note: Box plots are also called 'box and whiskers plots'.