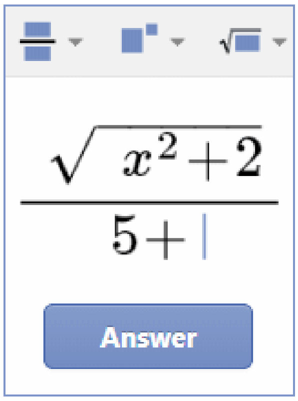
Index of lessons | Purplemath's lessons in offline form |
Forums | Print this page (print-friendly version) | Find local tutors

Search



Box-and-Whisker Plots: Interquartile Ranges and Outliers *(page 3 of 3)*

Sections: Quartiles, boxes, and whiskers, Five-number summary, Interquartile ranges and outliers

The "interquartile range", abbreviated "IQR", is just the width of the box in the box-and-whisker plot. That is, $\mathrm{IQR} = Q_3 - Q_1$. The IQR can be used as a measure of how spread-out the values are. Statistics assumes that your values are clustered around some central value. The IQR tells how spread out the "middle" values are; it can also be used to tell when some of the other values are "too far" from the central value. These "too far away" points are called "outliers", because they "lie outside" the range in which we expect them.

Box-and-Whisker Plots

The IQR is the length of the box in your box-and-whisker plot. An outlier is any value that lies more than one and a half times the length of the box from either end of the box. That is, if a data point is below $Q_1 - 1.5\times\mathrm{IQR}$ or above $Q_3 + 1.5\times\mathrm{IQR}$, it is viewed as being too far from the central values to be reasonable. Maybe you bumped the weigh-scale when you were making that one measurement, or maybe your lab partner is an idiot and you should never have let him touch any of the equipment. Who knows? But whatever their cause, the outliers are those points that don't seem to "fit".

(Why one and a half times the width of the box? Why does that *particular* value demark the difference between "acceptable" and "unacceptable" values? Because, when John Tukey was inventing the box-and-whisker plot in 1977 to display these values, he picked $1.5\times\mathrm{IQR}$ as the demarkation line for outliers. This has worked well, so we've continued using that value ever since.)

- **Find the outliers, if any, for the following data set:**

  10.2, 14.1, 14.4. 14.4, 14.4, 14.5, 14.5, 14.6, 14.7, 14.7, 14.7, 14.9, 15.1, 15.9, 16.4

  To find out if there are any outliers, I first have to find the IQR. There are fifteen data points, so the median will be at position $(15 + 1) \div 2 = 8$. Then $Q_2 = 14.6$. There are seven data points on either side of the median, so Q_1 is the fourth value in the list and Q_3 is the twelfth: $Q_1 = 14.4$ and $Q_3 = 14.9$. Then $\mathrm{IQR} = 14.9 - 14.4 = 0.5$.

  Outliers will be any points below $Q_1 - 1.5\times\mathrm{IQR} = 14.4 - 0.75 = 13.65$ or above $Q_3 + 1.5\times\mathrm{IQR} = 14.9 + 0.75 = 15.65$.

  Then the outliers are at 10.2, 15.9, and 16.4.

The values for $Q_1 - 1.5\times\mathrm{IQR}$ and $Q_3 + 1.5\times\mathrm{IQR}$ are the "fences" that mark off the "reasonable" values from the outlier values. Outliers lie outside the fences.

If your assignment is having you consider

MATHHELP LESSONS

Search

Our lessons matched to your textbook

Elementary / Middle / High School Math

5th Grade Math
6th Grade Math
Pre-Algebra
Algebra 1
Geometry
Algebra 2

Community College / College Math

College Pre-Algebra
Introductory Algebra
Intermediate Algebra
College Algebra

Standardized Test Prep

SAT Math
ACT Math
GED Math
ACCUPLACER Math
COMPASS Math
CLEP Math
GRE Math
GMAT Math
ASVAB Math
PRAXIS Math
TEAS Math
HESI Math
TABE Math

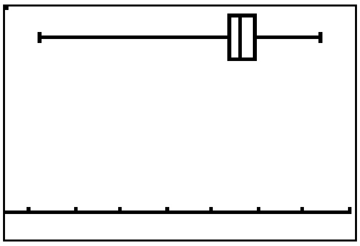
more tests...

outliers and "extreme values", then the values for $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$ are the "inner" fences and the values for $Q_1 - 3 \times IQR$ and $Q_3 + 3 \times IQR$ are the "outer" fences. The outliers (marked with asterisks or open dots) are between the inner and outer fences, and the extreme values (marked with whichever symbol you didn't use for the outliers) are outside the outer fences.

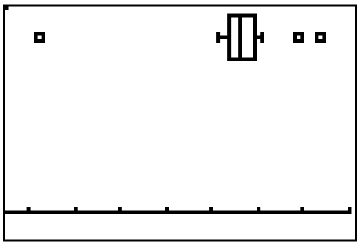
By the way, your book may refer to the value of "$1.5 \times IQR$" as being a "step". Then the outliers will be the numbers that are between one and two steps from the hinges, and extreme value will be the numbers that are more than two steps from the hinges.

Looking again at the previous example, the outer fences would be at $14.4 - 3 \times 0.5 = 12.9$ and $14.9 + 3 \times 0.5 = 16.4$. Since 16.4 is right on the upper outer fence, this would be considered to be only an outlier, not an extreme value. But 10.2 is fully below the lower outer fence, so 10.2 would be an extreme value.

Your graphing calculator may or may not indicate whether a box-and-whisker plot includes outliers. For instance, the above problem includes the points 10.2, 15.9, and 16.4 as outliers. One setting on my graphing calculator gives the simple box-and-whisker plot which uses only the five-number summary, so the furthest outliers are shown as being the endpoints of the whiskers:



A different calculator setting gives the box-and-whisker plot with the outliers specially marked (in this case, with a simulation of an open dot), and the whiskers going only as far as the highest and lowest values that aren't outliers:



Note that my calculator makes no distinction between outliers and extreme values.

If you're using your graphing calculator to help with these plots, make sure you know which setting you're supposed to be using and what the results mean, or the calculator may give you a perfectly correct but "wrong" answer.

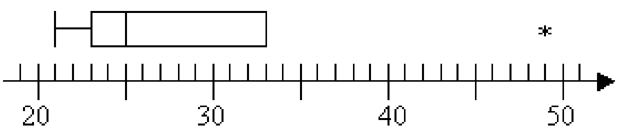
- **Find the outliers and extreme values, if any, for the following data set, and draw the box-and-whisker plot. Mark any outliers with an asterisk and any extreme values with an open dot.**

  $21, \ 23, \ 24, \ 25, \ 29, \ 33, \ 49$

  To find the outliers and extreme values, I first have to find the IQR. Since there are seven values in the list, the median is the fourth value, so $Q_2 = 25$. The first half of the list is $21, 23, 24$, so $Q_1 = 23$; the second half is $29, 33, 49$, so $Q_3 = 33$. Then $IQR = 33 - 23 = 10$.

  The outliers will be any values below $23 - 1.5 \times 10 = 23 - 15 = 8$ or above $33 + 1.5 \times 10 = 33 + 15 = 48$. The extreme values will be those below $23 - 3 \times 10 = 23 - 30 = -7$ or above $33 + 3 \times 10 = 33 + 30 = 63$.

  So **I have an outlier at 49 but no extreme values**, I won't have a top whisker because Q_3 is also the highest non-outlier, and my plot looks like this:

It should be noted that the methods, terms, and rules outlined above are what I have taught and what I have most commonly seen taught. However, your course may have different specific rules, or your calculator may do computations slightly differently. You may need to be somewhat flexible in finding the answers specific to your curriculum.

<< Previous Top | 1 | 2 | 3 | Return to Index

Cite this article as: Stapel, Elizabeth. "Box-and-Whisker Plots: Interquartile Ranges and Outliers." Purplemath. Available from http://www.purplemath.com/modules/boxwhisk3.htm. Accessed 03 February 2016