

Outliers and Modified Boxplots

by Sophia



WHAT'S COVERED

This tutorial will cover the topic of outliers and modified boxplots. Our discussion breaks down as follows:

1. Outliers
2. The 1.5xIQR Rule
3. Boxplots

1. Outliers

You may recall that **outliers** are values that are far outside the pattern established by the rest of the data. They're either very high or very low in comparison to the rest of the data set.

Boxplots, introduced in another tutorial, are a way to graphically display the five number summary for a data set. This tutorial will present a modified version of boxplots so that it is easier to observe outliers in them.

🔗 **EXAMPLE** Here is a set of test scores.

90, 98, 89, 88, 46, 90, 91, 84, 94

Almost everyone scored in the 80's or 90's, except for one student, who scored a 46. That student is an outlier.



TERMS TO KNOW

Outlier

A point that is so large or small as to be unusual, given the rest of the data points.

Modified Boxplot

A graphical display showing a modified version of the five number summary. If a distribution has outliers, then the "whiskers" only extend to the highest and lowest points that are not outliers.

2. The 1.5xIQR Rule

So, how do you use the 1.5xIQR method?



⇒ **EXAMPLE** Consider the data set of test scores from above.

Page 2



TRY IT

Suppose that you have the data set of all house prices for homes purchased in Albuquerque, New Mexico, from February to April in 1993. These are in thousands of dollars.

The first and third quartiles have been calculated:

Home Prices in Albuquerque, New Mexico From February - April, 1993					
205	72	93.9	99.5	87.5	105
208	72	82	97.5	88.9	104.5
215	74.9	78	97.5	85.5	105
215	73.1	77	90	83.5	102
199.9	72.5	70	96	81	100
190	67	62	86	80.5	103
180	215	54	169.5	79.9	97.5
156	159.9	107	155.3	75	95
145	135	210	125	75.9	94
144.9	129.9	72.5	130	75.5	92
137.5	125	66	102	75	94.5
127	123.9	60	102	73	87.4
125	120	58	92.2	72.9	87.2
123.5	112.5	184.4	92.5	71	87
117	110	158	89.9	77.3	86.9
118	108	69.9	85	69	76.6
115.5	105	133	87.6	67	73.9
111	104.9	116	89	61.9	
113.9	95.5	110.9	87	129.5	
99.5	93.4	112.9	70	97.5	
Q1 = 78, Q3 = 120					

What is the range for outliers? What is the lower fence for outliers and the upper fence for outliers?

$$IQR = Q3 - Q1 = 120 - 78 = 42$$

1.5 IQR below the first quartile:

$$Q1 - 1.5IQR = 78 - (1.5)(42) = 78 - 63 = 15$$

1.5 IQR above the third quartile:

$$Q3 + 1.5IQR = 120 + (1.5)(42) = 120 + 63 = 183$$

The interquartile range is 42, so any point below 78 minus 1.5xIQR, which is 15, or above 120 plus 1.5xIQR, which is 183, will be an outlier.

Notice that there's nothing in the list below 15, but there are seven above 183. This means that there are seven outliers in this data set, which, by the way, is a completely legitimate and legal occurrence in this situation.



TERM TO KNOW

1.5xIQR Rule

If a point is larger than $Q3 + 1.5xIQR$, or smaller than $Q1 - 1.5xIQR$, then it is an outlier.

3. Boxplots

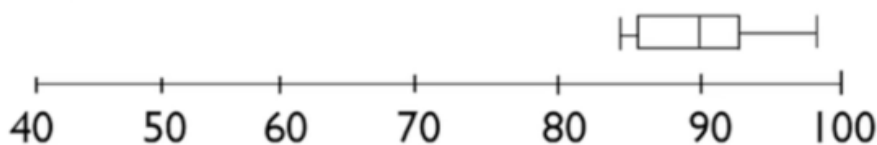
You can use this new information to create a new version of an already existing plot that you have. You've made boxplots in another tutorial; now you can modify them to show outliers.

Generally, you would make the whiskers on the box-and-whisker plot extend all the way out to the maximum and minimum. If the minimum or maximum (or both) are outliers, that will make the whiskers really long. For a modified boxplot, instead of going all the way out to those outliers, you can extend them only to the highest and lowest values that aren't outliers and notate the outliers separately.

⇒ **EXAMPLE** Refer back to the student data set from the section above. Here are the values from least to greatest.

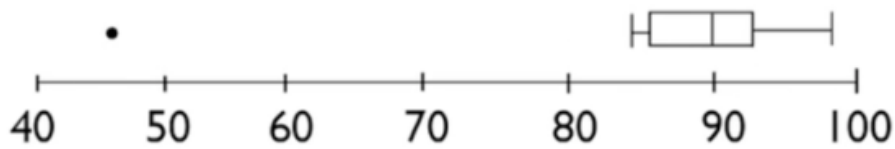


Mark the same values that you would have if you were making a regular box-and-whisker plot. However, don't go all the way down to 46 for your minimum--even though 46 is the actual minimum. 46 is an outlier, so instead go to the next lowest number that isn't an outlier--84--and make your line there. Then you can make your box and whiskers.

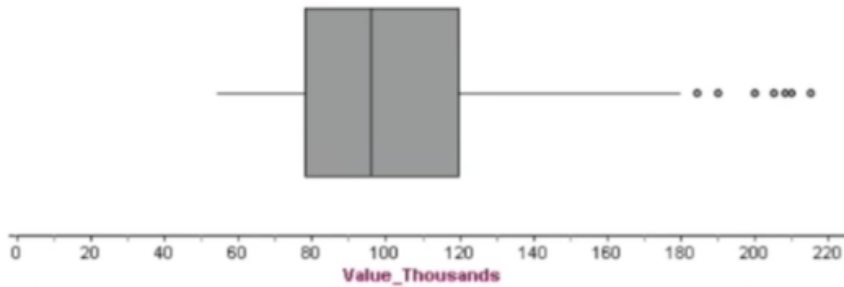


You still have to show the 46 as part of this data set somehow, so you will mark it with a dot. This is a

modified boxplot.



In the home value data set, there were seven high outliers. This is a modified plot for that data set:



SUMMARY

You can determine in some measurable way if a point within a data set is an outlier using the $1.5 \times \text{IQR}$ rule. Data sets might have no outliers, or they might have one or more outliers on the low side, one or more outliers on the high side, or both. There's no rule for how many outliers are allowed in a data set. Whatever outliers exist, you can use a modified boxplot to visually display them.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR [TERMS OF USE](#).



TERMS TO KNOW

1.5xIQR Rule

If a point is larger than $Q3 + 1.5 \times \text{IQR}$, or smaller than $Q1 - 1.5 \times \text{IQR}$, then it is an outlier.

Modified Boxplot

A graphical display showing a modified version of the five number summary. If a distribution has outliers, then the "whiskers" only extend to the highest and lowest points that are not outliers.

Outlier

A point that is so large or small as to be unusual, given the rest of the data points.