Using the IQR to Detect Outliers

Learning Objective: Summarize and describe the distribution of a quantitative variable in context: a) describe the overall pattern, b) describe striking deviations from the pattern.

Learning Objective: Relate measures of center and spread to the shape of the distribution, and choose the appropriate measures in different contexts.

Using the IQR to Detect Outliers

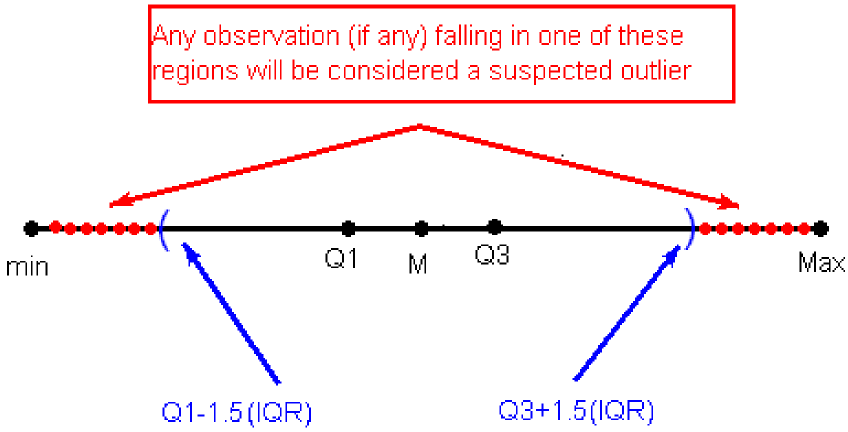
So far we have quantified the idea of center, and we are in the middle of the discussion about measuring spread, but we haven't really talked about a method or rule that will help us classify extreme observations as outliers. The IQR is used as the basis for a rule of thumb for identifying outliers.

The 1.5(IQR) Criterion for Outliers

An observation is considered a suspected outlier if it is:

- below Q1 - 1.5(IQR) or
- above Q3 + 1.5(IQR)

The following picture illustrates this rule:

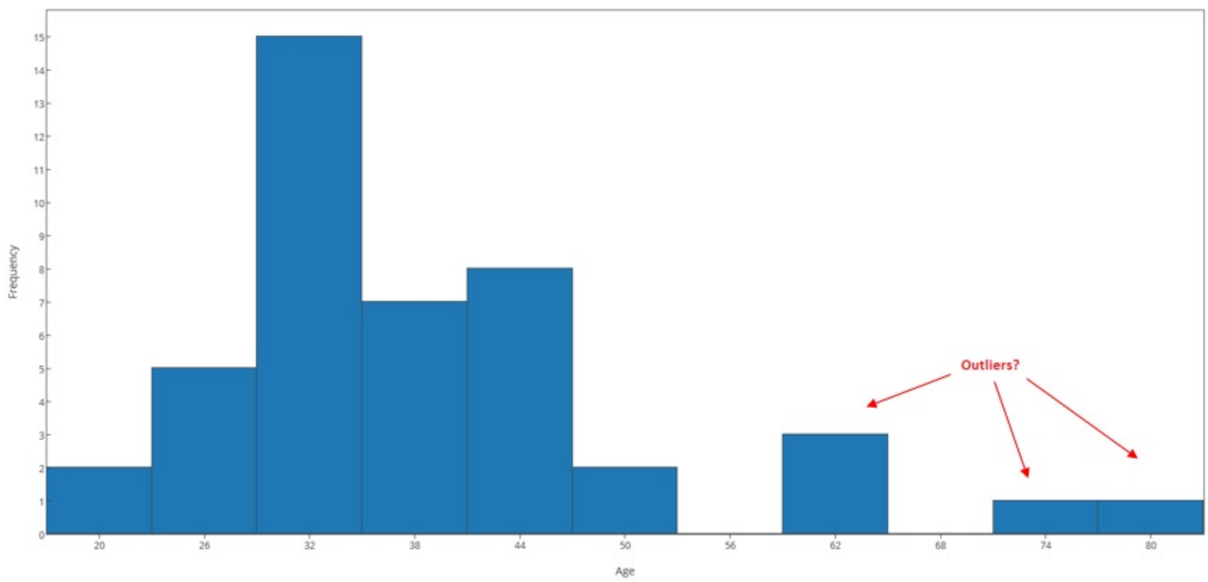


Example: Best Actress Oscar Winners

We will continue with the Best Actress Oscar winners example (To see the full dataset, click here.)

34 34 27 37 42 41 36 32 41 33 31 74 33 49 38 61 21 41 26 80 42 29 33 36 45 49 39 34 26 25 33 35 35 28 30 29 61 32 33 45 29 62 22 44

Recall that when we first looked at the histogram of ages of Best Actress Oscar winners, there were five observations that looked like possible outliers:



We can now use the 1.5(IQR) criterion to check whether the five observations should indeed be classified as outliers:

- For this example we found that $Q1 = 30.5$ and $Q3 = 42 \Rightarrow \mathrm{IQR} = 11.5$
- $Q1 - 1.5(\mathrm{IQR}) = 30.5 - (1.5)(11.5) = 13.25$

- $Q3 + 1.5\,(\mathrm{IQR}) = 42 + (1.5)\,(11.5) = 59.25$

The 1.5(IQR) criterion tells us that any observation that is below 13.25 or above 59.25 is considered a suspected outlier.

We therefore conclude that the observations 61, 61, 62, 74 and 80 should be flagged as suspected outliers in the distribution of ages. Note that since the smallest observation is 21, there are no suspected low outliers in this distribution.

Did I Get This

1/1 point (graded)
A survey taken in a large statistics class contained the question: "What's the fastest you have driven a car (in miles per hour)?" The five-number summary for the 87 males surveyed is:

min = 55, Q1 = 95, Median = 110, Q3 = 120, Max = 155

Should the largest observation in this data set be classified as an outlier?

- ○ Yes
- ◉ No ✔

Answer
Correct:
The IQR in this case is 120 - 95 = 25. Applying the 1.5(IQR) rule, we find: Q3 + 1.5(IQR) = 120 + 1.5(25) = 157.5, and therefore the largest observation, 155, should NOT be classified as an outlier. Note, however that in this case: Q1 - 1.5(IQR) = 95 - 1.5(25) = 57.5, and therefore the smallest observation, 55, should be classified as an outlier.

Submit

Did I Get This

1/1 point (graded)
A survey taken of 140 sports fans asked the question: "What is the most you have ever spent for a ticket to a sporting event?"

The five-number summary for the data collected is:

min = 85, Q1 = 130, Median = 145, Q3 = 150, Max = 250

Should the smallest observation be classified as an outlier?

- (selected) Yes ✔
- No

Answer

Correct:

The IQR is 150 - 130 = 20. Using the 1.5(IQR) criterion we get 130 - 1.5(20) = 100. Since the smallest observation of 85 is smaller than 100, it should be considered an outlier.

Submit