The Big Earth Data Project

# Lesson 2: Exploring outliers

The following rules are often used for outliers:

**Using quartiles:**

* More than 1.5 inter-quartile ranges above the upper quartile or below the lower quartile.

**Using mean and standard deviation:**

* More than 2 standard deviations away from the mean.
* More than 3 standard deviations away from the mean.

### Identifying outliers

Show that the minimum value for this data could be considered an outlier.

|  |  |
| --- | --- |
| September ozone hole area (million km2) | |
| Count | 44 |
| Mean | 16.67 |
| Std | 6.61 |
| Min | 0.14 |
| 25% | 14.77 |
| 50% | 18.74 |
| 75% | 21.72 |
| Max | 24.73 |

**Questions**

Determine whether there are outliers in any of these sets of data for the ozone hole area for October, November or December.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| October ozone hole area (million km2) | |  | November ozone hole area (million km2) | |  | December ozone hole area (million km2) | |
| Count | 44 |  | Count | 44 |  | Count | 44 |
| Mean | 16.13 |  | Mean | 8.66 |  | Mean | 1.39 |
| Std | 5.52 |  | Std | 6.16 |  | Std | 2.11 |
| Min | 1.70 |  | Min | 0.00 |  | Min | 0.00 |
| 25% | 13.97 |  | 25% | 3.10 |  | 25% | 0.00 |
| 50% | 17.09 |  | 50% | 8.56 |  | 50% | 0.08 |
| 75% | 20.43 |  | 75% | 14.55 |  | 75% | 2.30 |
| Max | 25.48 |  | Max | 19.16 |  | Max | 8.70 |

Classifying outliers

**Ozone hole area (from the monthly data)**  
A graph of a graph showing different colored squares

Description automatically generated with medium confidence

A graph of a graph with lines

Description automatically generated with medium confidence

State whether you believe that the value of 8.7 million square kilometres in December 2020 should be removed from the data. Give reasons for your answer.

**Stratospheric temperature in January (from the daily data)**

|  |  |  |  |
| --- | --- | --- | --- |
| Show that there is at least one outlier in this data set  State whether you believe that the outlier you have found should be removed from the data and give reasons for your answer. |  | January temperatures in the stratosphere (°C) | |
| Count | 44 |
| Mean | -48.67 |
| Std | 1.96 |
| Min | -55.29 |
| 25% | -49.67 |
| 50% | -48.59 |
| 75% | -47.66 |
| Max | 0.00 |