

## Outlier detection explained

- E.g. on a DNA microarray you have multiple probes per gene and each probe is applied multiple times per array
- How to combine multiple values to one value
- 10, 11, 14, 16, 18, 18, 19
  - Mean = 15.1
  - Median = 16 ( sort and take the middle )
- 10, 11, 14, 16, 18, 18, 59
  - Mean = 20.9
  - Median = 16 ← This seems better
- 10, 11, 11, 11, 18, 18, 19
  - Mean = 14 ← This seems better
  - Median = 11
- Uniform solution: remove outliers
  - $Q1 - 1.5 * IQR$  and  $Q3 + 1.5 * IQR$
  - $IQR = Q3 - Q1$
  - Uniform solution: remove outliers
  - $Q1 - 1.5 * IQR$  and  $Q3 + 1.5 * IQR$
  - $IQR = Q3 - Q1$
- Calculation: 10, 11, 11, 11, 18, 18, 19
  - $N = 7$  ← total numbers
  - Sort the numbers
  - $Q1 = \text{round}(N/4) == \text{round}(7/4) = 2 \rightarrow$  the second value is here 11
  - $Q3 = N - Q1 + 1 == 7 - 2 + 1 = 6 \rightarrow$  sixth number is 18
  - $IQR = Q3 - Q1 == 18 - 11 = 7$
  - $Q1 - 1.5 * IQR == 11 - 1.5 * 7 = 11 - 10.5 = 0.5 \rightarrow$  remove everything  $< 0.5$
  - $Q3 + 1.5 * IQR == 18 + 1.5 * 7 = 18 + 10.5 = 28.5 \rightarrow$  remove everything  $> 28.8$
- 10, 11, 14, 16, 18, 18, 59
  - $Q1 - 1.5 * IQR == 11 - 1.5 * 7 = 11 - 10.5 = 0.5 \rightarrow$  remove everything  $< 0.5$
  - $Q3 + 1.5 * IQR == 18 + 1.5 * 7 = 18 + 10.5 = 28.5 \rightarrow$  remove everything  $> 28.8$
  - Mean = 14.5
  - Median = 15
  - Both seems better now!!