Percentiles

The $p$th percentile of a set of measurements has at least $p\%$ of the measurements at or below it and at least $(100-p)\%$ of the measurements at or above it.
Example

Consider the data set 2, 3, 3, 4, 8

Percentiles

SAS has five ways of calculating percentiles!

Quartiles

Consider the data set 2, 3, 3, 4, 8

Inter-quartile Range

IQR = 75\textsuperscript{th} percentile – 25\textsuperscript{th} percentile

“Five number summary” is the minimum, 25\textsuperscript{th} percentile, median, 75\textsuperscript{th} percentile, and the maximum.

Which To Use?

Data is approximately symmetric and unimodal – use the mean and standard deviation

Otherwise – use the five number summary
Example revisited

Consider the data set 2, 3, 3, 4, 8

Outliers

Outliers are values that are unusual in the context of the data set. If the data consists of one variable they are usually the values that are unusually large or unusually small.
Common explanations of outliers are:
1) Error in observing or recording the value
2) Comes from a different population
3) A rare event

Outliers

Outliers can only be removed in the case where it is clearly an error in observation or recording… not just because you think it was an error.

One option is to report the results both with and without the outlier(s).