# **CONCRETE COMPRESSIVE STRENGTH DATASET ANALYSIS**

Summary of the columns in the dataset:

cement slag ash water superplastic

Min. :102.0 Min. : 0.0 Min. : 0.00 Min. :121.8 Min. : 0.000

1st Qu.:192.4 1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.:164.9 1st Qu.: 0.000

Median :272.9 Median : 22.0 Median : 0.00 Median :185.0 Median : 6.400

Mean :281.2 Mean : 73.9 Mean : 54.19 Mean :181.6 Mean : 6.205

3rd Qu.:350.0 3rd Qu.:142.9 3rd Qu.:118.30 3rd Qu.:192.0 3rd Qu.:10.200

Max. :540.0 Max. :359.4 Max. :200.10 Max. :247.0 Max. :32.200

coarseagg fineagg age strength

Min. : 801.0 Min. :594.0 Min. : 1.00 Min. : 2.33

1st Qu.: 932.0 1st Qu.:731.0 1st Qu.: 7.00 1st Qu.:23.71

Median : 968.0 Median :779.5 Median : 28.00 Median :34.45

Mean : 972.9 Mean :773.6 Mean : 45.66 Mean :35.82

3rd Qu.:1029.4 3rd Qu.:824.0 3rd Qu.: 56.00 3rd Qu.:46.13

Max. :1145.0 Max. :992.6 Max. :365.00 Max. :82.60

Observations from summary: (use box plots for outliers)

Data doesn’t seem to have any missing values.

Cement - Mean and median are close, indicating a normal distribution

Slag – Median is much lesser than mean, indicating a long right tail.

Ash – Median is 0, indicating many concrete compositions do not have ash added, also has long right tail

Water – More or less normal distribution

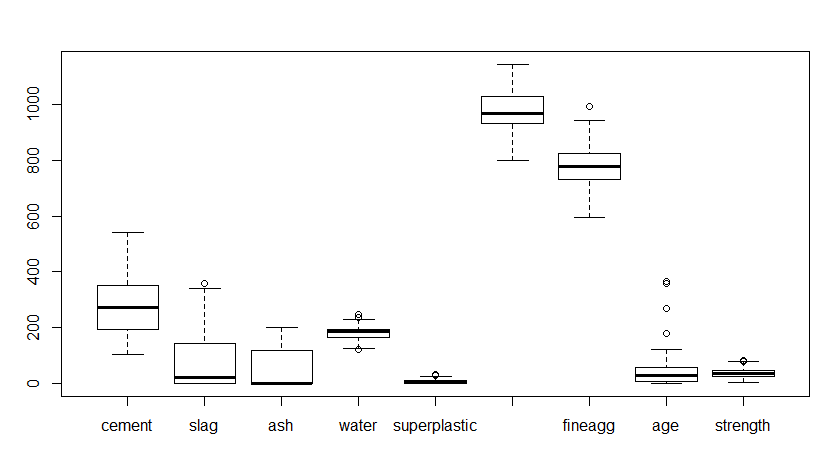
Superplastic – Median and mean are almost equal, but 1st quartile is zero , data is skewed towards the right

Coarse\_Agg – More or less normal distribution

Fine\_Agg – More or less normal

Age – Samples are tested for strength at different periods in a year (1~365) ,

Strength = normal



Observations from the boxplot:

Most of the samples seem to have been tested in the early stages (AGE), few outliers in the column

Ash and slag are heavily skewed towards the right, many compositions have 0 ash and slag

Coarse, fine aggregate, cement are mostly normal distributions

Q1, Median, Q3 lie very close to each other for super plastic, with a long right whisker and few outliers –

the spread of data is very tight, values are close to each other.

Strength of the concrete is also normal distribution, with few outliers

Water also has outliers , but not by much.