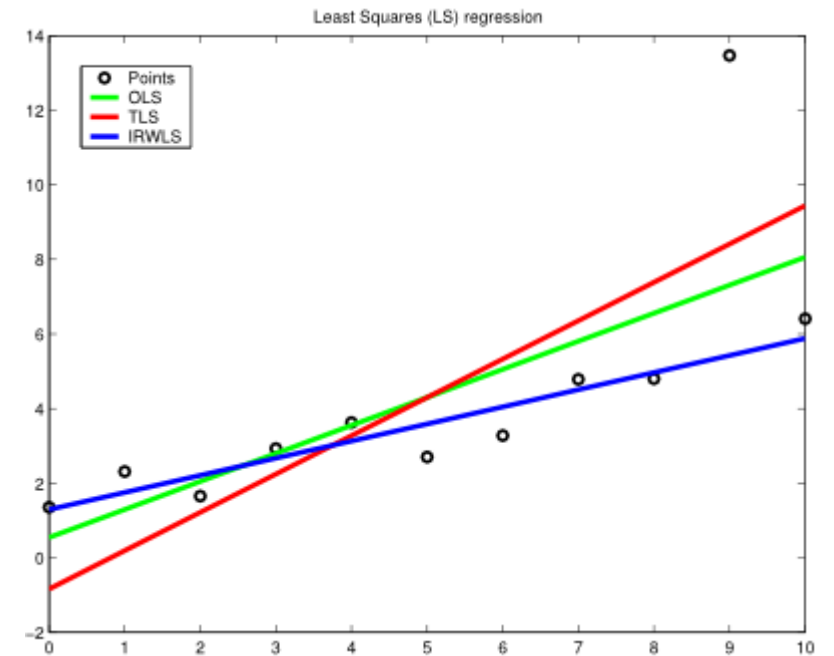


# Robust Statistics

Outliers



# Robust statistics

What do you do if your error distribution is not Gaussian?

This is also called a non-normal distribution. For example, if your data contain many outliers, they will not be well represented by a Gaussian.

Robust statistics was developed for this type of data

The median is a robust measure of central tendency, while the mean is not; for instance, the median has a breakdown point of 50%, while the mean has a breakdown point of 0% (a single large sample can throw it off).

The median absolute deviation (mad) and interquartile range are robust measures of statistical dispersion, while the standard deviation and range are not.

# Median and mad

The median is the numerical value separating the higher half of a data sample, a population, or a probability distribution from the lower half. The *median* of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one.

The median absolute deviation (mad) is the median of the difference between each numerical value and the median.

$$\text{mad} = \text{median} ( |X - \text{median}| )$$

For a normal distribution the median is essentially the same as the mean (or average). However, the mad differs from the standard deviation by a factor of 1.483. This value is the inverse of the normal inverse “cumulative distribution function” valued at  $\frac{3}{4}$ . The cumulative distribution function  $F(x)$  is the probability that  $x$  is less than  $X$ ,  $P(x < X)$ .