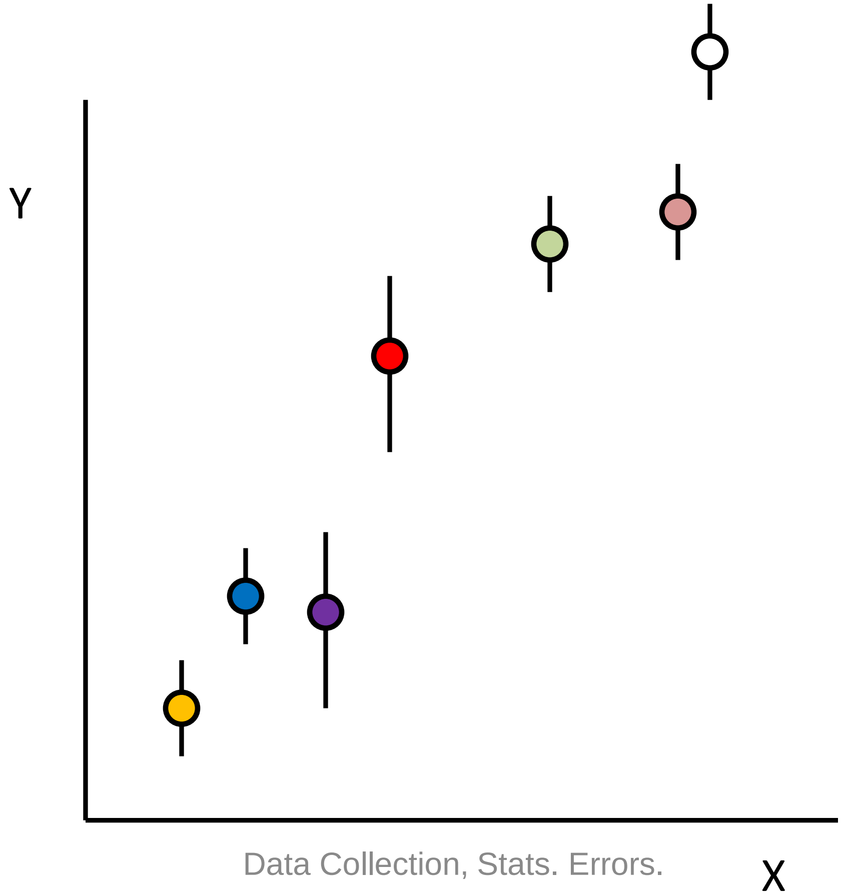
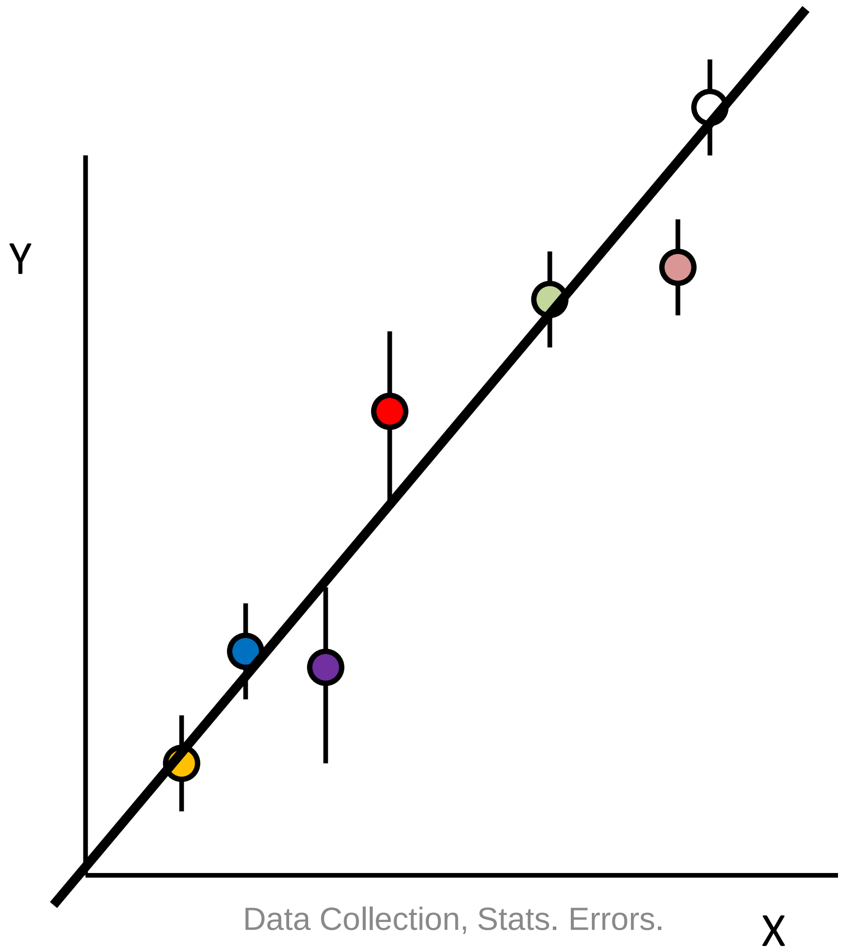
Estimating Parameter Errors with the bootstrap method

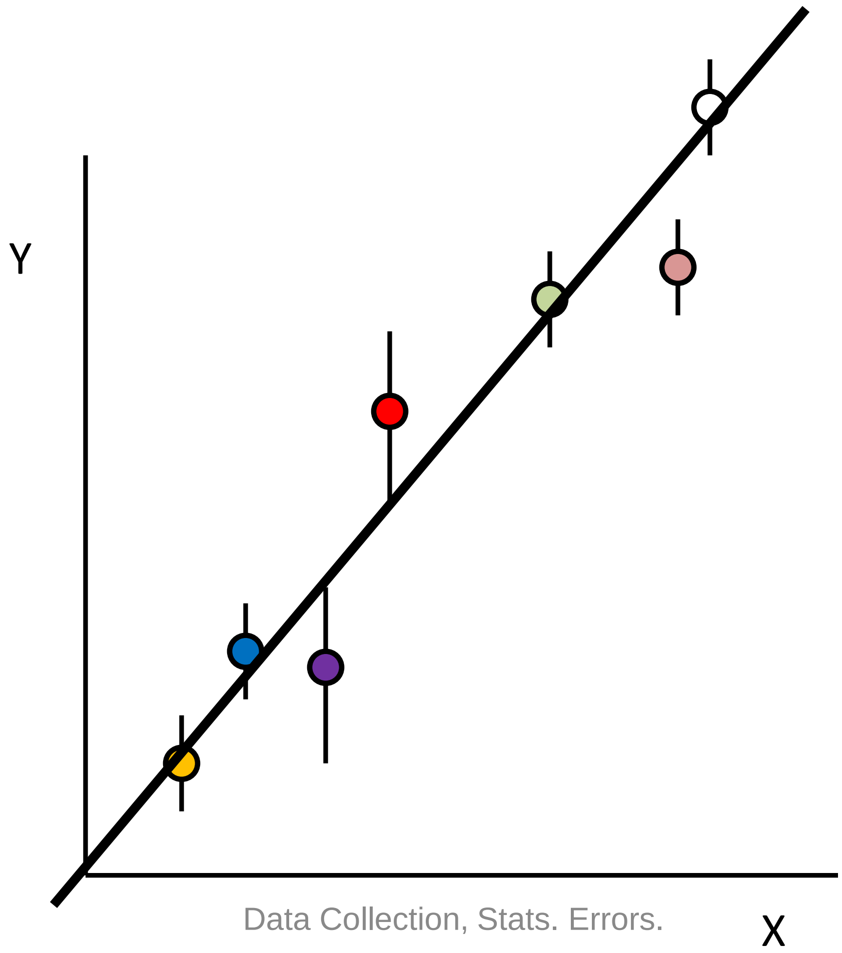
Suppose you have a set of measurements as shown below.
X and Y correspond to two variables.
For example X could be change in time and Y could be change in position in a measurement of the speed of light.



You can fit a line through these points described as y=mx+b where you find the values for m and b that yield the best fit. The fit could be a Chi-Squared minimization, for example



You have a limited sample of data.
How do you estimate the uncertainty the fit y=mx+b ?
ie what are the errors in m an b given the errors in the measured points.
One method is called the "Boot Strap Method"



Bootstrap method: An example

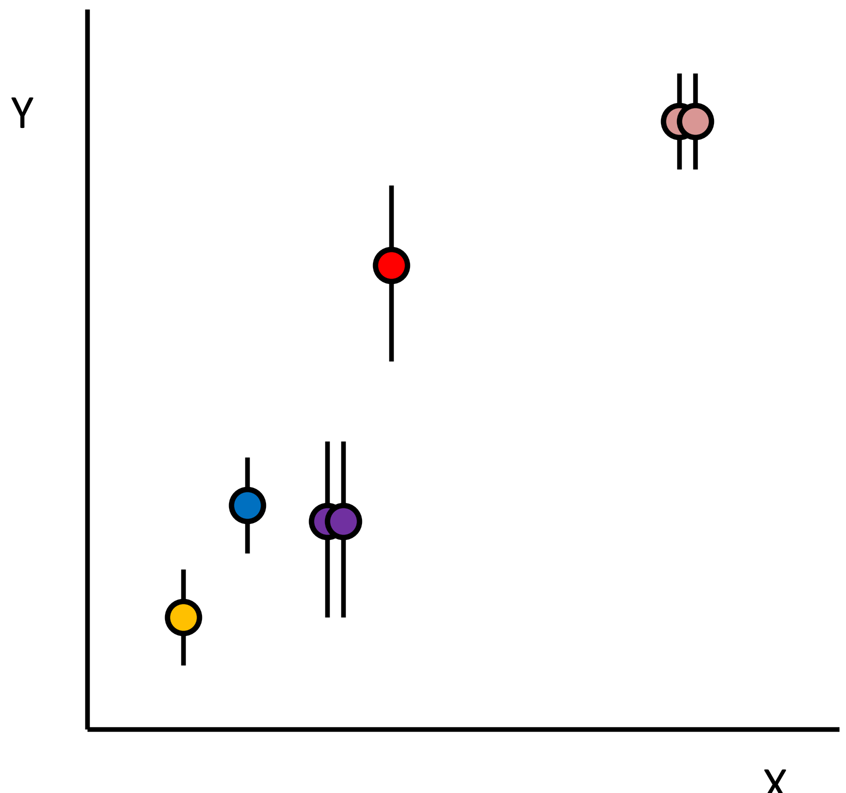
Put the 7 data points in a "bag".

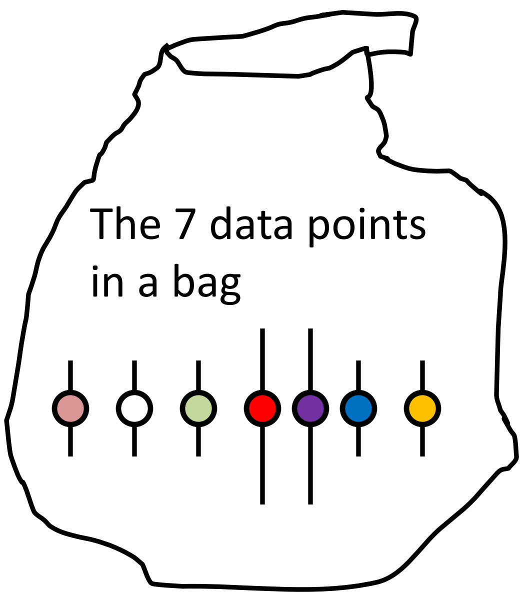
Pick one at random, plot it, and put it back in the bag

Do this 7 times. You now have 7 points on your graph.

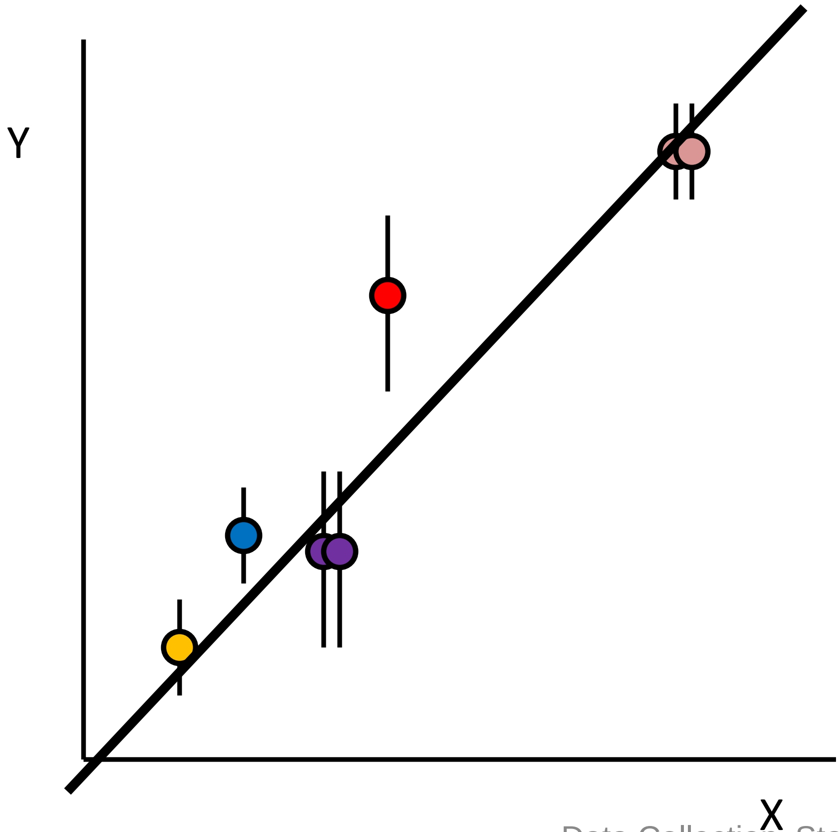
But some points may be picked several times and others not at all.

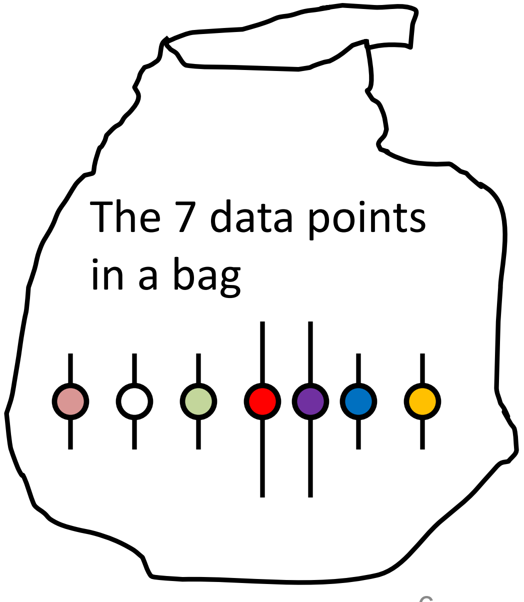
In this example below, two points were picked twice. Two points were not picked at all (The pairs of points were offset artificially on this plot for illustration)



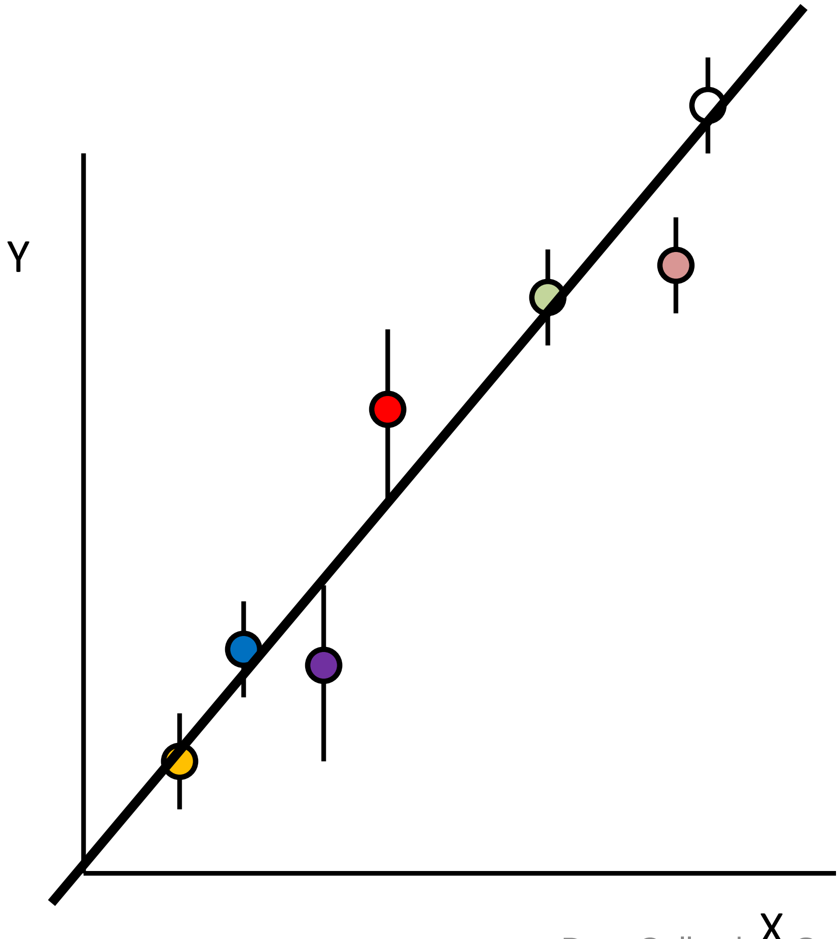


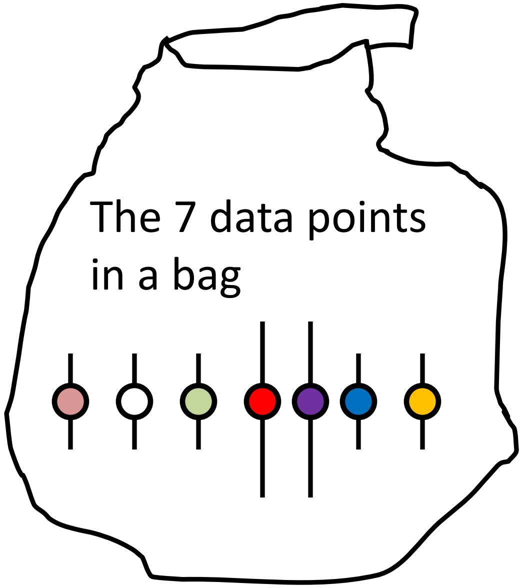
Then fit the this sampled data set to y=mx+b
Note that the line of best fit will probably be different because the data set is probably different (unlikely you'll select each point exactly once)



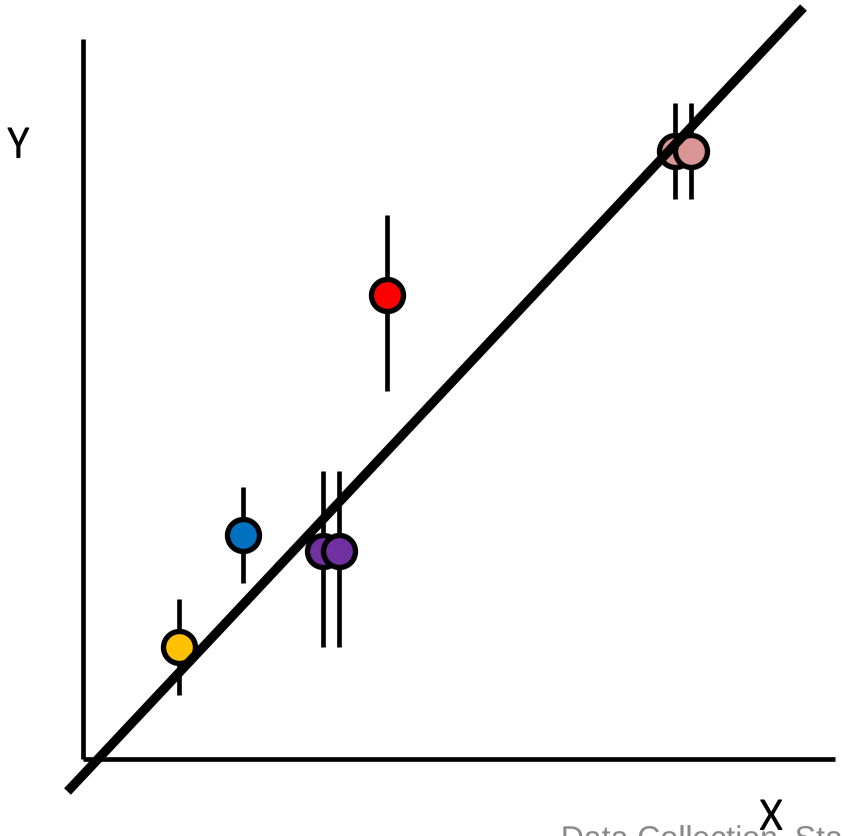


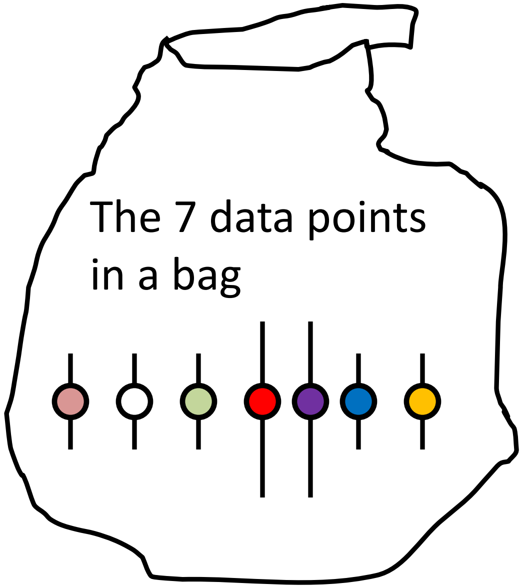
Data Sample that was collected (shown for reference)





Then save the value you got for m and b (the two parameters in the fit)



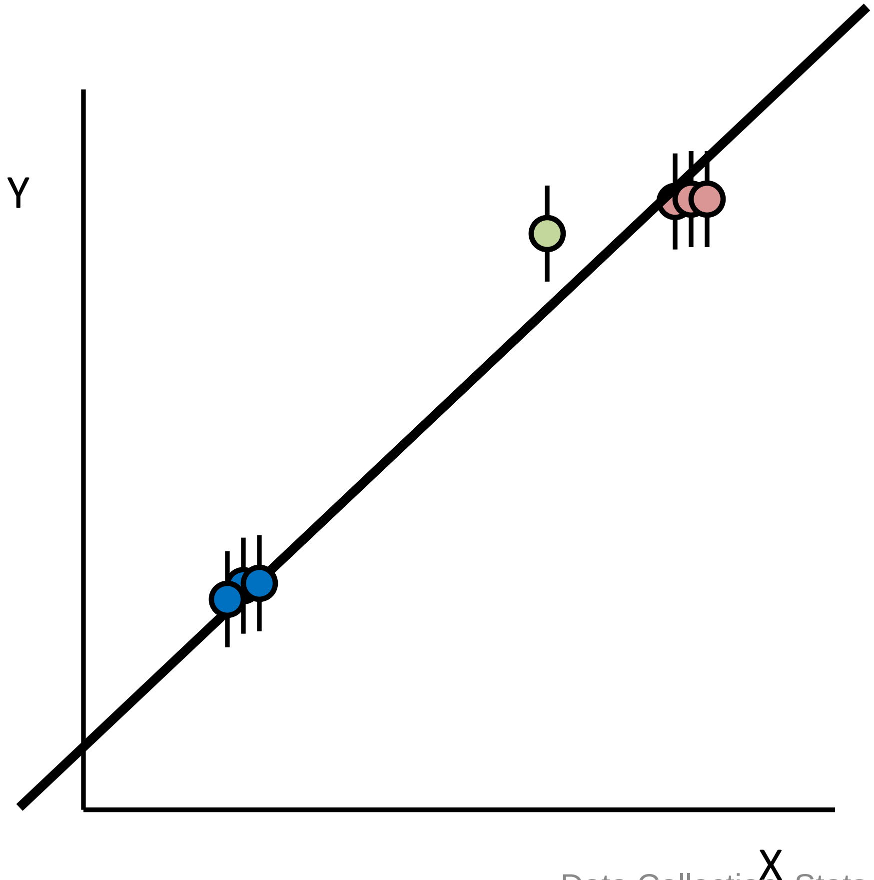


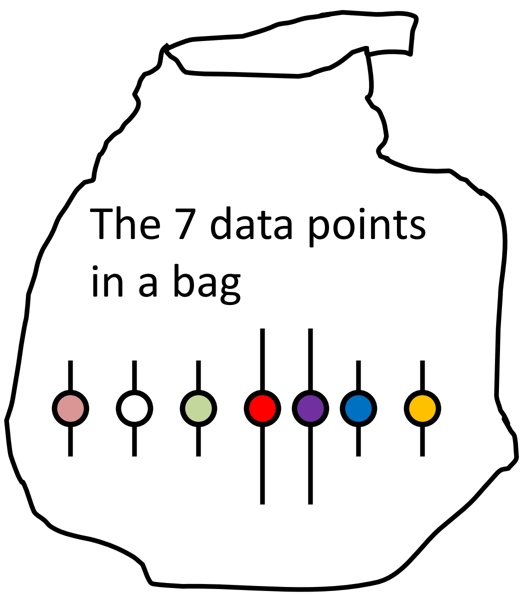
Then iterate this process about 100 times

100 resampled data sets

100 fits (one from each data set)

100 pairs of values for m and for b





So now you have 100 values
for m (m1, m2, m3, …… m100)
for b (b1, b2, b3, …… b100)

Calculate m_{rms} of these 100 m numbers
Calculate the b_{rms} of these 100 b numbers

rms root mean square

m_{rms}=1/100 $\Sigma(\langle m\rangle - m_i)^2$

$\langle m\rangle$ is 1/100 Σm_i

The RMS of m_{rms} and b_{rms} are the statistical errors in m and b.

You can also histogram the 100 m numbers and 100 b numbers and examine the distributions

For a detailed discussion see

Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy

Author(s): B. Efron and R. Tibshirani

Source:

Statistical Science,

Vol. 1, No. 1 (Feb., 1986), pp. 54-75

Note this paper has >4000 citations