Big Data Analytics

Four “V” of Big Data:  
- volume (data size)  
- velocity(speed of change)  
- variety(data types)  
- veracity(bias, noise, anomaly)

Standard deviation - how data is spread out:  
Sx = sqrt(sum(xi - mean) over (n-1)) [unbiased: n-1]  
Example:  
x = 3,1,3,9 mean = 4 sqrt((3-4)^2 + (1-4)^2 + (3-4)^2 + (9-4)^2 / n or n- 1)= sqrt(1+9+1+25)/3 = 3.46 (unbiased)

Covarience - dimensional quantity to check relationships between x and y:  
Sxy = sum(xi-mean)*(yi-mean)/n-1*  
*Example:*  
*x = 3, 1, 3, 9*  
*x mean = 4*  
*y = 5, 2, 4, 1*  
*y mean= 3*  
*(3-4)*(5-3)+… = -10/4 = -2.5

Bayes Rule:  
P(H|E) = P(E|H) \* P(H) / P(E)