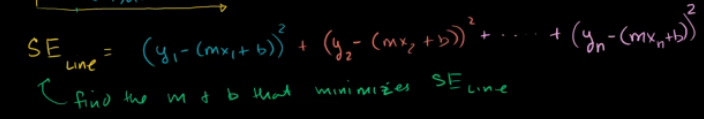
**统计学笔记4**

# 线性回归

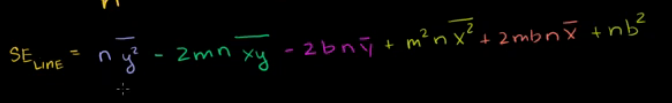
问题核心：寻找能够具有最佳拟合效果的直线（y =mx+b）来表示一个样本集的整体趋势

Minimize 总方差来求解最佳的m和b

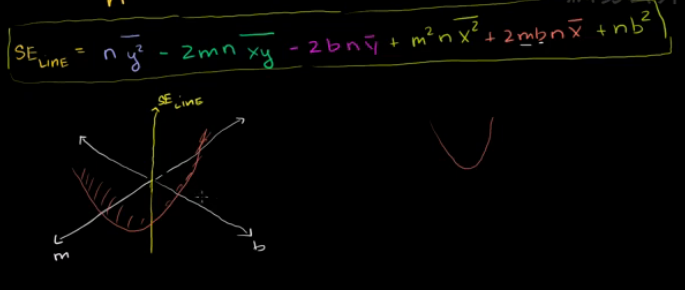
总方差表示为



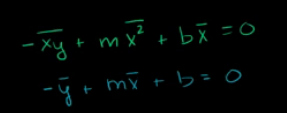
通过化简合并得到



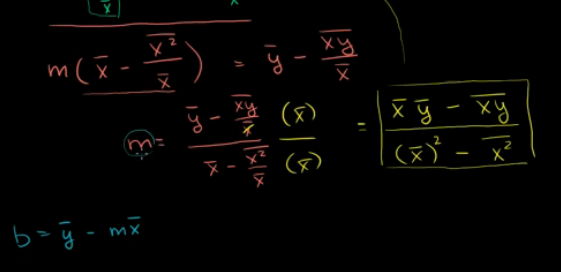
表示所有y值的均值，表示所有x值的均值，n表示样本点的数量  
得到的总方差方程可以看作是由m，b，方差组成的曲面



分别对m,b求偏导数就再令求出的方程为0，就可求得最优解

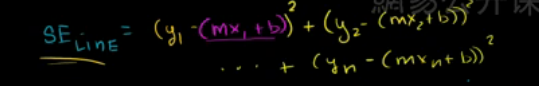


最终的算法为

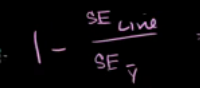


# R2

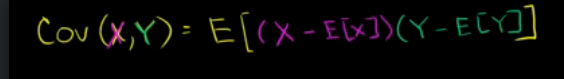
含义：y的总波动的百分之多少是可以由x的波动所描述的？



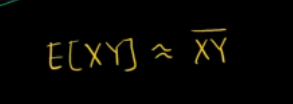




# 协方差



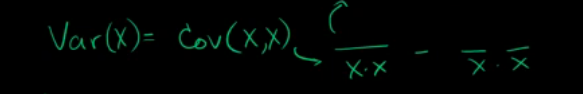
两个变量上多大程度上一同变化



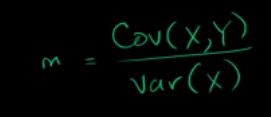
最后得到



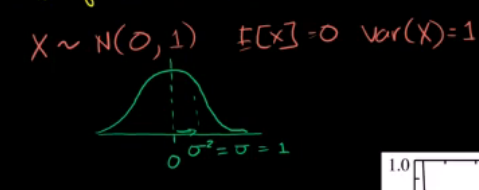
变量自身的协方差等于自身方差



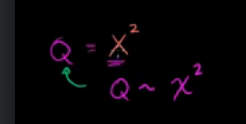
跟线性回归结合



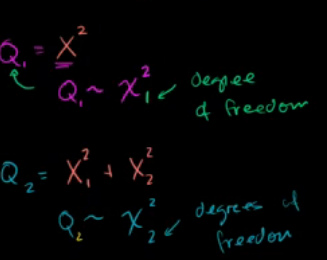
# 卡方分布



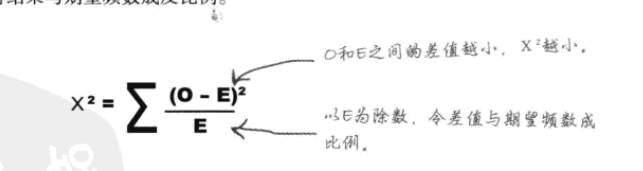
Q表示从标准正态分布中抽样的x的值的平方



组成自由度为1的卡方分布



检验统计量卡方分布提供一种观察频数与期望频数之间的差异进行度量的算法



# 方差分析

