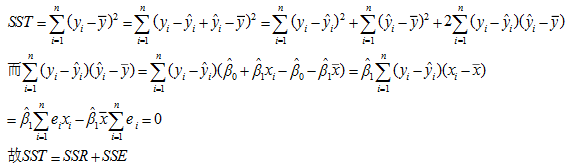
1. 线性回归

使用直线去拟合给定的n个点，线性回归的任务是最小化平方误差，即求，通过对该式变形化简，然后分别对m和b求偏导，得到以下结果：

()和()这两点在拟合的直线上。

线性回归模型不仅适用于二维，也适用于多维。

决定系数R2反应了y的波动有多少百分比能被x的波动所描述，，其中，称为残差平方和；SSR=，称为总平方和。SST=SSR+SSE，证明过程为：



1. 协方差

协方差的意义：度量各个维度偏离其均值的程度。

协方差的值如果为正，则说明两者是正相关的(从协方差可以引出“相关系数”的定义)，结果为负值就说明负相关的，如果为0，也是就是统计上说的“相互独立”。

协方差计算公式：

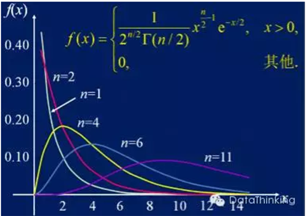
随机变量与本身的协方差即为该随机变量的方差，即：

线性回归中，回归线的斜率

1. 卡方分布

设X1,X2,......Xn相互独立, 都服从标准正态分布N(0,1), 则称随机变量χ2=X12+X22+......+Xn2所服从的分布为自由度为 n 的χ2分布.

卡方分布的期望E（χ2）=n，方差D(χ2)=2n。



1. 卡方检验

卡方检验是一种用途很广的计数资料的假设检验方法。它属于非参数检验的范畴，主要是比较两个及两个以上样本率( 构成比）以及两个分类变量的关联性分析。其根本思想就是在于比较理论频数和实际频数的吻合程度或拟合优度问题。

卡方检验的计算公式为：

其中，A为实际值，T为理论值。

x2用于衡量实际值与理论值的差异程度（也就是卡方检验的核心思想），包含了以下两个信息：

1. 实际值与理论值偏差的绝对大小（由于平方的存在，差异是被放大的）。

2. 差异程度与理论值的相对大小。

列联表卡方检验中自由度n=(row-1)\*(col-1)。

1. 方差分析

，称为总平方和；SSW=，m为组数，SSW称为组内平方和；SSB=，称为组间平方和。SST=SSW+SSB。

F统计量越大，说明总体波动更多来自于组间波动，越小说明总体波动更多来自组内波动。

1. 归纳推理与演绎推理

归纳推理（inductive reasoning）：需要找规律或趋势，然后推广。就是用已有信息进行趋势外推。归纳推理在推广时，并不知道趋势是否会继续，只是假设它会继续。

演绎推理(deductive reasoning)：从既有的事实或数据出发，演绎得到其他正确的事实，演绎推理知道肯定正确。