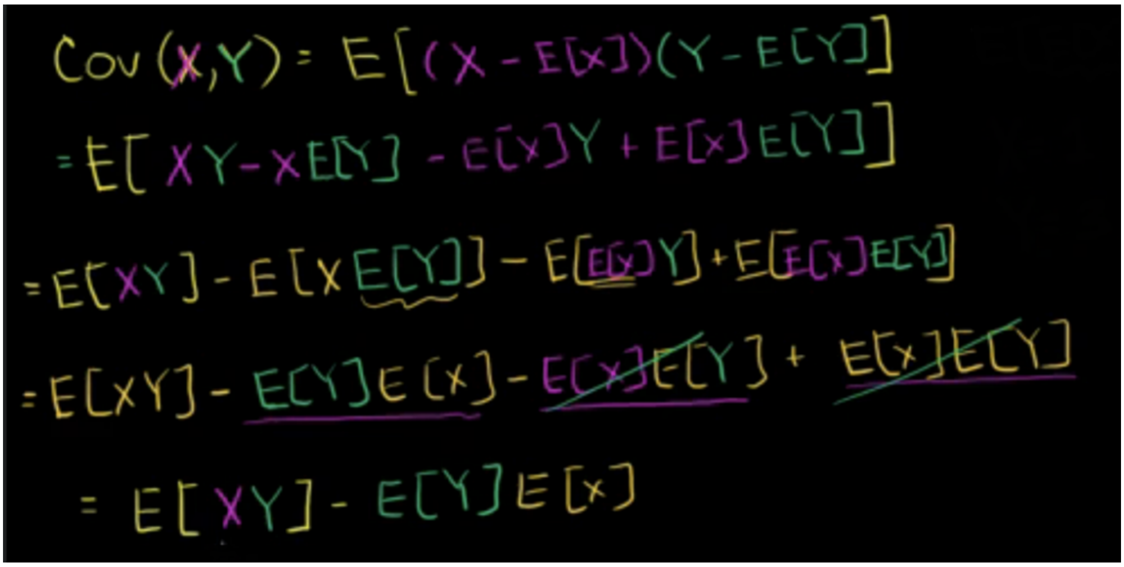
协方差（covariance）： 两随机变量离各自均值距离之积的期望值

$$\mathrm{Cov}({\color{blue}X},{\color{red}Y}) = E[({\color{blue}X - E(X)}) \cdot ({\color{red}Y - E(Y)})]$$

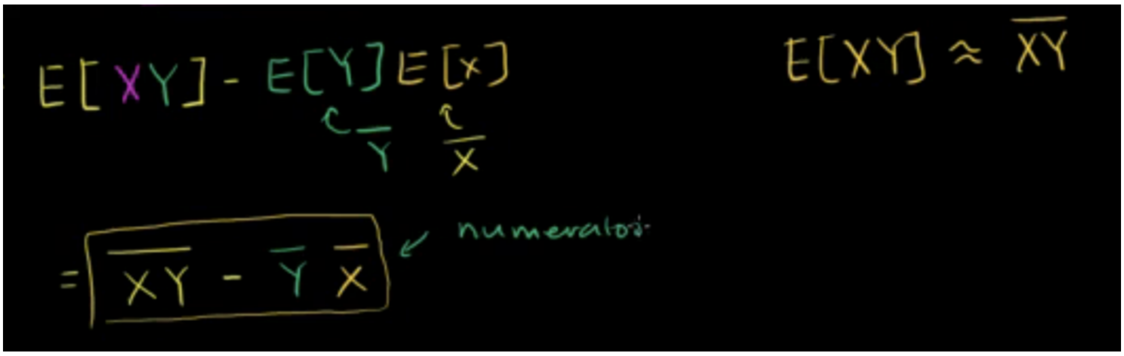
协方差也就是表示两变量多大程度上一同变化。如果两个变量的实际值均低于或高于其均值，则是正协方差。否则是负协方差。

我们这里要研究的是协方差与回归线的关系。

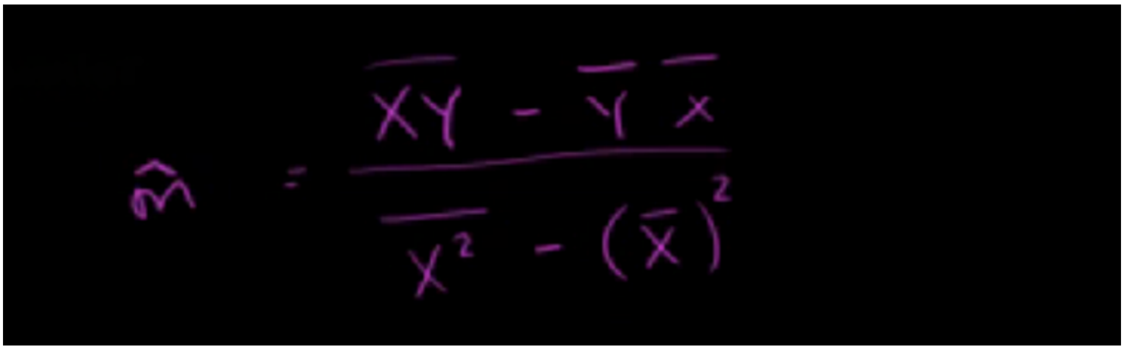
首先，重写协方差的公式：



而XY的期望值可以近似为XY相乘后所有乘积的均值。X的期望值可以近似为X所有样本的均值、Y的期望值可以近似为Y的所有样本的均值。上面的公式可以改写为：



而这就是回归线斜率的分子部分。我们来看下回归线斜率：



分母部分可以改写为：

$$\overline{X \cdot X} - \bar{X}\bar{X}$$

也就是X与X的协方差：

$$\mathrm{Cov}(X, X)$$

变量与自身的协方差即变量的方差：

$$\mathrm{Var}(X)$$

所以回归线斜率就是X,Y的协方差除以X的方差。

$$m = \frac{\mathrm{Conv}(X, Y)}{\mathrm{Var}(X)}$$