

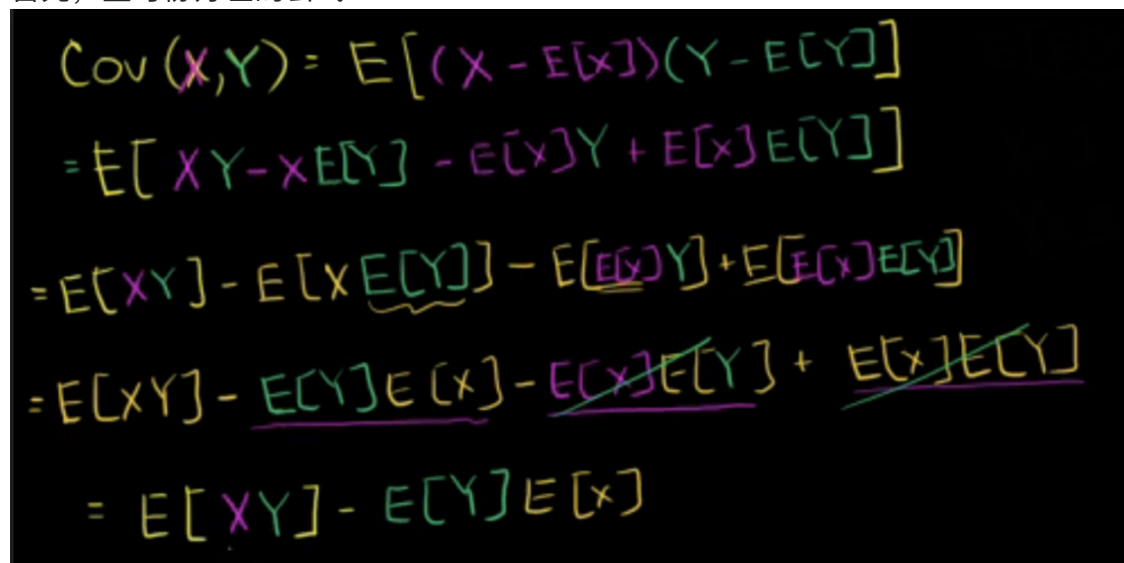
协方差 (covariance) : 两随机变量离各自均值距离之积的期望值

$$\text{Cov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

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。

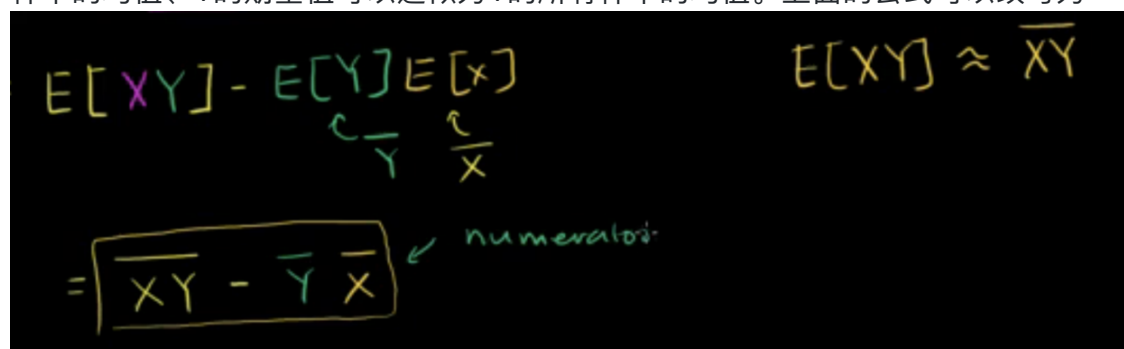
我们这里要研究的是协方差与回归线的关系。

首先，重写协方差的公式：



A handwritten derivation of the covariance formula on a black background. The steps are as follows:
1. $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$
2. $= E[XY - XE[Y] - E[X]Y + E[X]E[Y]]$
3. $= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]]$
4. $= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y]$ (Note: $E[XE[Y]] = E[Y]E[X]$ and $E[E[X]Y] = E[X]E[Y]$)
5. $= E[XY] - E[Y]E[X]$

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：



A handwritten simplified formula for covariance on a black background. It shows:
1. $E[XY] - E[Y]E[X]$ with $E[Y]$ labeled as \bar{Y} and $E[X]$ labeled as \bar{X} .
2. $E[XY] \approx \overline{XY}$
3. $= \boxed{\overline{XY} - \bar{Y}\bar{X}}$ with an arrow pointing to the box and the word "numerator".

而这就是回归线斜率的分子部分。我们来看下回归线斜率：

$$\hat{m} = \frac{\overline{XY} - \bar{Y}\bar{X}}{\overline{X^2} - (\bar{X})^2}$$

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：

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

变量与自身的协方差即变量的方差：

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

所以回归线斜率就是X,Y的协方差除以X的方差。

$$m = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$