

协方差和相关系数

Guangyao Zhao

2022-08-13

Contents

协方差的性质	2
相关系数的性质	2
相关 or 独立	2

在数字特征之方差的时候有一个假设说, 当 X, Y 相互独立的时候: $D(X + Y) = D(X) + D(Y)$, 否则 $D(X + Y) = D(X) + D(Y) - 2E\{[X - E(X)] \cdot [Y - E(Y)]\}$ 。为了研究两个变量不独立的情况, 即两变量会存在某种关系, 所以就有必要研究下 $E\{[X - E(X)] \cdot [Y - E(Y)]\}$ 。统计学家将这一指标定义为协方差 (Covariance)

协方差的定义如下:

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

上式有点类似方差的公式: $D(X) = E(X^2) - E^2(X)$

协方差的主要目的是评价 X, Y 的关系, 进一步来讲是线性关系。但是该指标有量纲, 比如变量 X 代表重量 (kg), 变量 Y 代表长度 (m), 那么协方差的单位为 $kg \cdot m$, 在根据某个指标描述相关关系的时候往往将其最大值设置为 1, 这样更容易建立出感性的认识, 所以引出了相关系数 (Correlation coefficient) 这一概念。其中皮尔逊相关系数最为出名:

$$\rho = \frac{Cov(X, Y)}{D(X)D(Y)}$$

协方差的性质

- $Cov(X, X) = D(X)$
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, c) = 0$
- $Cov(aX, bY) = abCov(X, Y)$
- $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

相关系数的性质

- $|\rho| \leq 1$
- $|\rho| = 1$ 的充要条件是存在常数 a, b , 使 $P\{Y = a + bX\} = 1$ 。也就是说当 X, Y 的样本点都在一条直线的时候 $|\rho| = 1$
- $|\rho| = 0$ 时 X, Y 相互独立, 即问题回到了最初

相关 or 独立

- 独立: 没有关系, 是完完全全不存在任何关系
- 不相关: 没有线性关系。相关不相关是在线性系统中对变量的描述, 范围更小

没有关系 (独立) 一定也没有线性关系, 但是没有线性关系不一定没有关系。