天津中德应用技术大学
TianjinSino-German University of Applied Sciences

第16讲：协方差与相关系数

天津中德应用技术大学 主讲：邱玉文 2022/11/15

本次内容概要

一、练习

二、协方差和相关系数

三、切比雪夫不等式

一、方差练习

一、方差的定义

1. 定义

$$D(X)=E\left\{\left[X-E(X)\right]^2\right\}$$

方差用来描述随机变量取值的波动（集中与分散）程度

2. 离散和连续情形

$$D(X)=\begin{cases}\sum_{i=1}^{\infty}(x_i-E(X))^2 p(x_i), & \text{离散情形}\\ \int_{-\infty}^{+\infty}(x-E(X))^2 f(x)dx, & \text{连续情形}\end{cases}$$

3. 计算*D*(*X*)的简便公式

$$D(X)=E(X^2)-\left[E(X)\right]^2$$

方差的定义 $D(X)=E\left\{[X-E(X)]^2\right\}$

计算方差的常用公式

$$D(X)=E(X^2)-[E(X)]^2.$$

由方差的定义及数学期望的性质得

$$\begin{aligned}D(X)&=E\{[X-E(X)]^2\}\\&=E\{X^2-2XE(X)+[E(X)]^2\}\\&=E(X^2)-2E(X)E(X)+[E(X)]^2\\&=E(X^2)-[E(X)]^2.\end{aligned}$$

3. 方差的性质

- 常数的方差为0，即 $D(c)=0$
- $D(aX+b)=a^2D(X)$
- 若随机变量X与Y相互独立，且方差都存在，则

$$D(X\pm Y)=D(X)+D(Y)$$
$$D(aX\pm bY)=a^2D(X)+b^2D(Y)$$

- 对任意随机变量X与Y，若它们的方差都存在，则

$$D(aX\pm bY)=a^2D(X)+b^2D(Y)\pm 2ab\operatorname{cov}(X,Y)$$

其中$\operatorname{cov}(X,Y)$称为X与Y的协方差

$$\operatorname{cov}(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$$

3、随机变量的标准化

1. 标准化变换公式

$$X^* = \frac{X - E(X)}{\sigma(X)}$$

X^* 称为X的标准化随机变量。

2. 标准化随机变量的均值和方差

$$E(X^*) = 0, \quad D(X^*) = 1$$

6、常用分布的期望和方差

分布名称及记号	概率函数或概率密度	数学期望	方差
"0—1"分布	$p(x)=p^x q^{1-x}, x=0,1.$ $(0<p<1, p+q=1)$	p	pq
二项分布 $B(n,p)$	$p(x)=C_n^x p^x q^{n-x}, x=0,1,\cdots,n.$ $(0<p<1, p+q=1)$	np	npq
超几何分布 $H(n,M,N)$	$p(x)=\dfrac{C_M^x C_{N-M}^{n-x}}{C_N^n}$, $x=0,1,\cdots,\min(n,M).$ $(0\leqslant n\leqslant N, 0\leqslant M\leqslant N)$	$\dfrac{nM}{N}$	$\dfrac{nM(N-M)(N-n)}{N^2(N-1)}$
泊松分布 $P(\lambda)$	$p(x)=\dfrac{\lambda^x e^{-\lambda}}{x!}, x=0,1,2,\cdots.$ $(\lambda>0)$	λ	λ

天津中德应用技术大学 TianjinSino-German University of Applied Sciences

6、常用分布的期望和方差（续）

泊松分布 $P(\lambda)$	$p(x)=\dfrac{\lambda^x e^{-\lambda}}{x!}, x=0,1,2,\cdots.$ $(\lambda>0)$	λ	λ
几何分布 $G(p)$	$p(x)=pq^{x-1}, x=1,2,3,\cdots.$ $(0<p<1, p+q=1)$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$
均匀分布 $U(a,b)$	$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b;\\ 0, & x<a \text{ 或 } x>b.\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布 $e(\lambda)$	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\leqslant 0.\end{cases}$ $(\lambda>0)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
正态分布 $N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)},$ $-\infty<x<+\infty.$ $(\sigma>0)$	μ	σ^2

四、练习题

49.随机变量X的概率密度为$f(x)=\begin{cases}Ax^3 & 0<x<a\\ 0 & 其它\end{cases}$，已知$E(X)=\dfrac{3}{5}$

试求：（1）a,A的值；（2）$P\{-\dfrac{1}{2}<X<\dfrac{1}{2}\}$；（3）分布函数$F(x)$.

$\boxed{E(X)=2}$ $\quad -2+3X$

1．已知随机变量 $X\sim B(10,0.2)$，设 $Y=3X-2$，则 $E(Y)=$（ 4 ）．

2．设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{x}{2}, & 0<x<2,\\ 0, & \text{其他},\end{cases}$$

$E(X)=\int_{-\infty}^{+\infty} xf(x)\,dx$

$=\int_0^2 \frac{1}{2} x^2\,dx$

$=\frac{1}{6}x^3\Big|_0^2$

$\frac{1}{6}\times 8$

则数学期望 $E(X)=$（ $\frac{4}{3}$ ）．

3．若随机变量 X 的概率函数为

X	-1	0	1
p	0.1	0.8	0.1

，则 $E(X^2)=$（ 0.2 ）．

$(-1)^2\times 0.1+(1)^2\times 0.1$

4．若二维离散型随机变量(X,Y)的联合概率分布为

$X \backslash Y$	1	2
-1	0.2	0.3
1	0.4	0.1

，

则$E(XY)=$（　　）．

$(-1)\times1\times0.2+ -1\times2\times0.3+1\times1\times0.4+1\times2\times0.1$

5．若随机变量X服从泊松分布$P(\lambda)$，已知$E(X)=2$，则$D(2X)=$（ 8 ）．

$D(X)=2$

$D(2X)=4D(X)$

$E(X)=\frac{6}{2}=3$

$\frac{(6-0)^2}{12}$

$D(X)=\frac{(6)^2}{12}=3$

6. 已知两个相互独立随机变量 $X\sim U(0,6), Y\sim e\left(\frac{1}{2}\right)$，则

$E(Y)=2$

$D(Y)=4$

$6-2$

$E(2X-Y)=$ （ 4 ），$D(2X-Y)=$ （ 8 ）；

$12-4$

7. 随机变量 $X\sim N(2,2^2), Y\sim B(4,0.5)$，则 $E(X^2+Y)=$ （ 6 ）

$E(Y)=2$　　$E(X)=2$

$4+$

二、协方差和相关系数

一、变量间的相关关系（不确定关系）

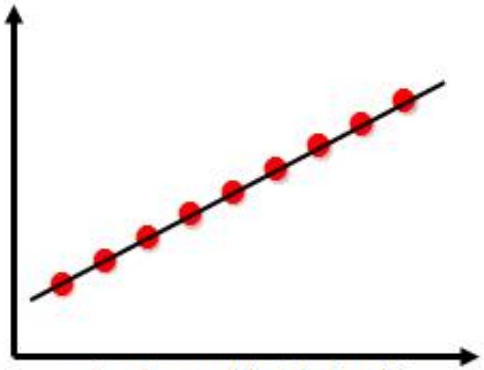
1. 变量间的相关关系

例如

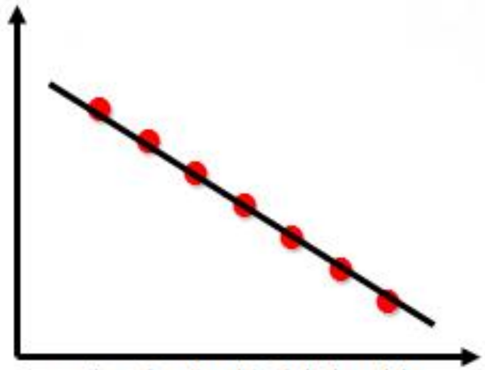
- 人的身高和体重
- 家庭的收入和消费
- 商品的广告费和销售额
- 粮食的施肥量和产量
- 股票的时间和价格
- 学生的期中和期末考试成绩, ···

$\boxed{X}$ $\boxed{Y}$

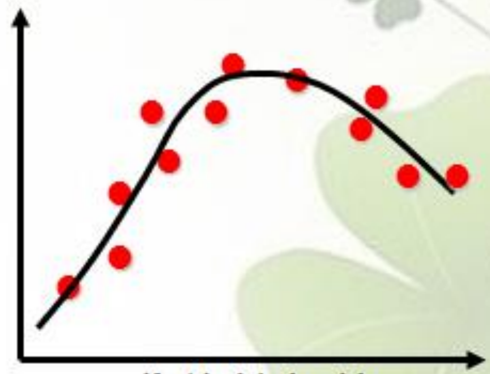
2．相关关系的图示



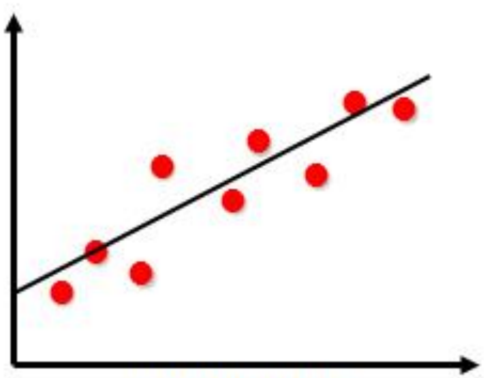
完全正线性相关



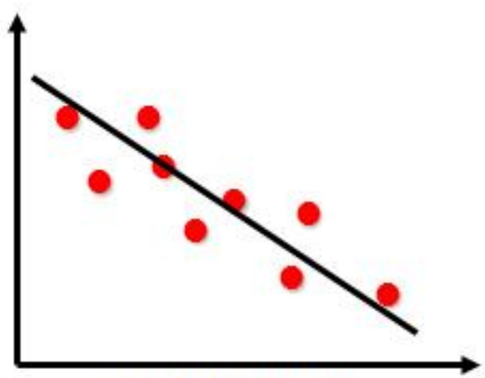
完全负线性相关



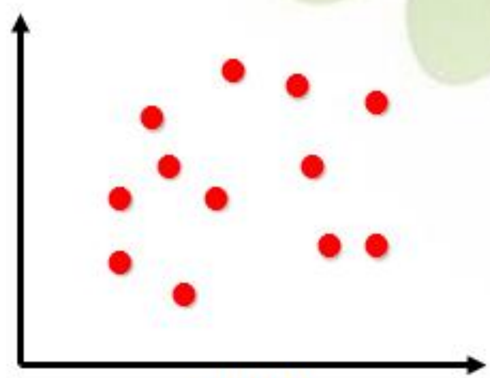
非线性相关



正线性相关

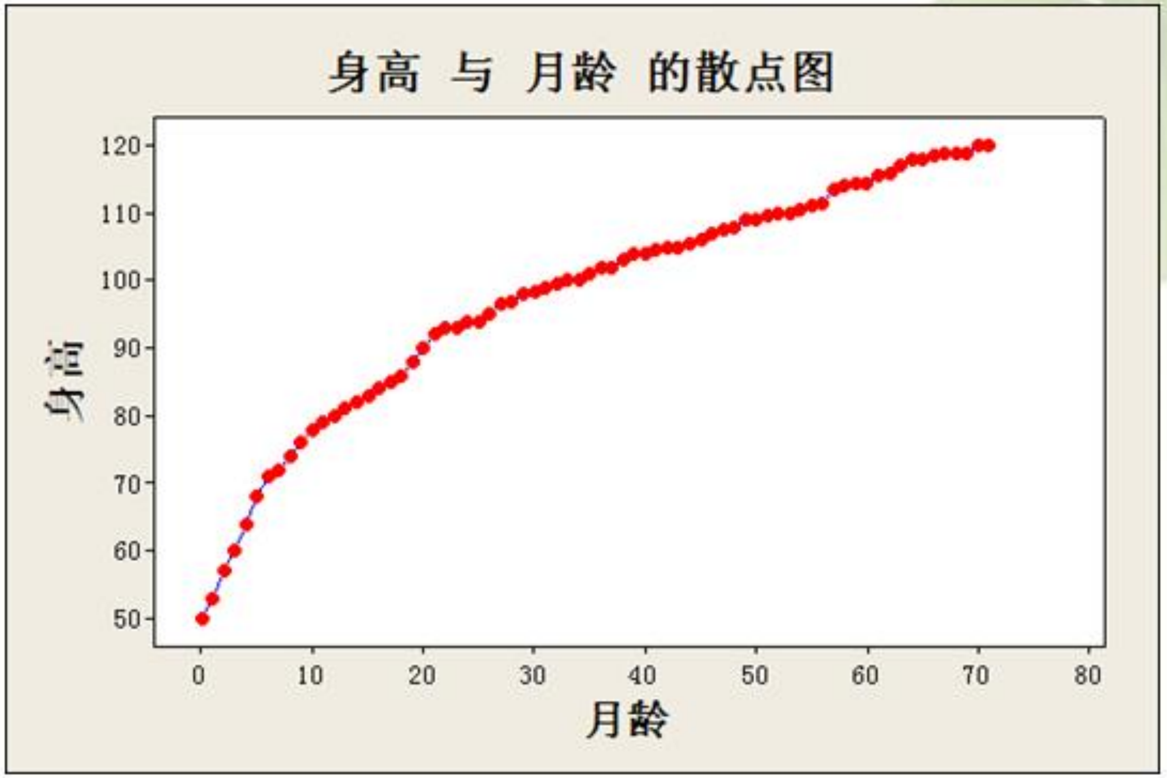


负线性相关

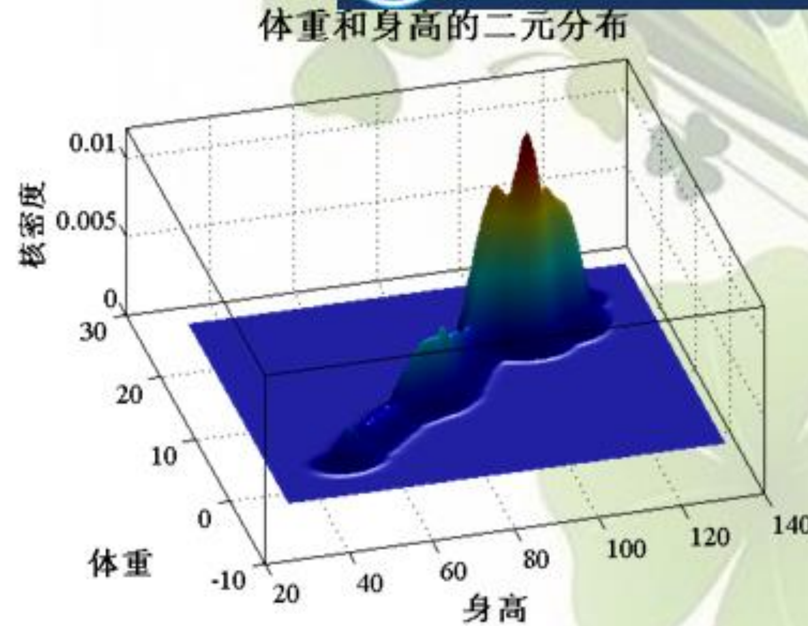


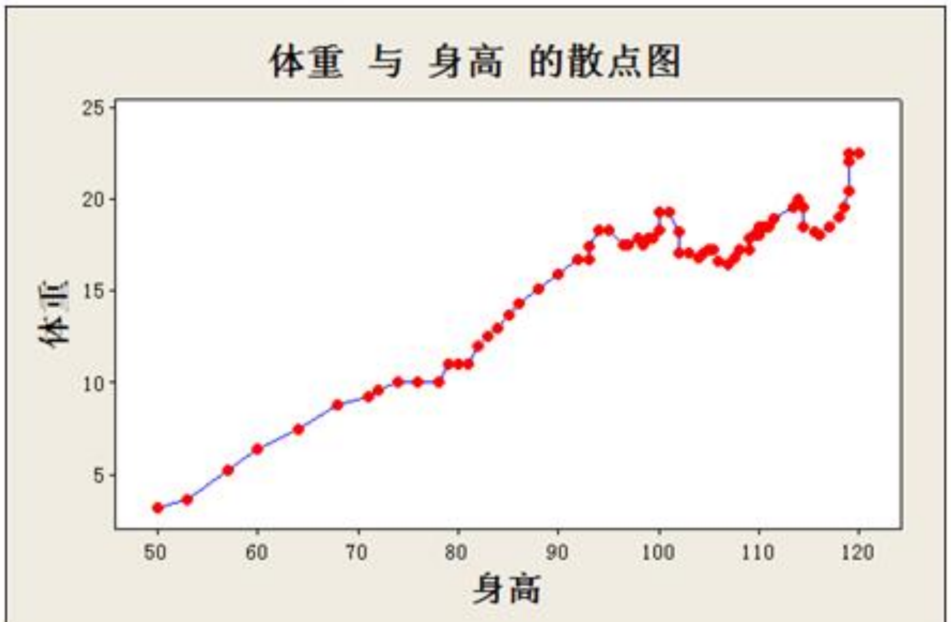
不相关

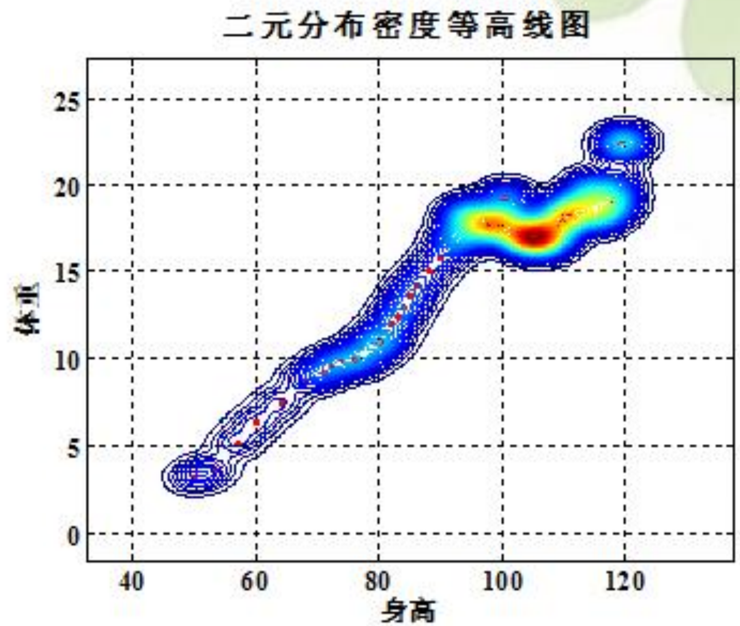
【例3.6-1】 有人从“雪林山庄——甜雨的开心乐园”中收集了一组儿童成长记录数据（0-7岁），包括月龄、身高和体重的观测数据。据此绘图如下：



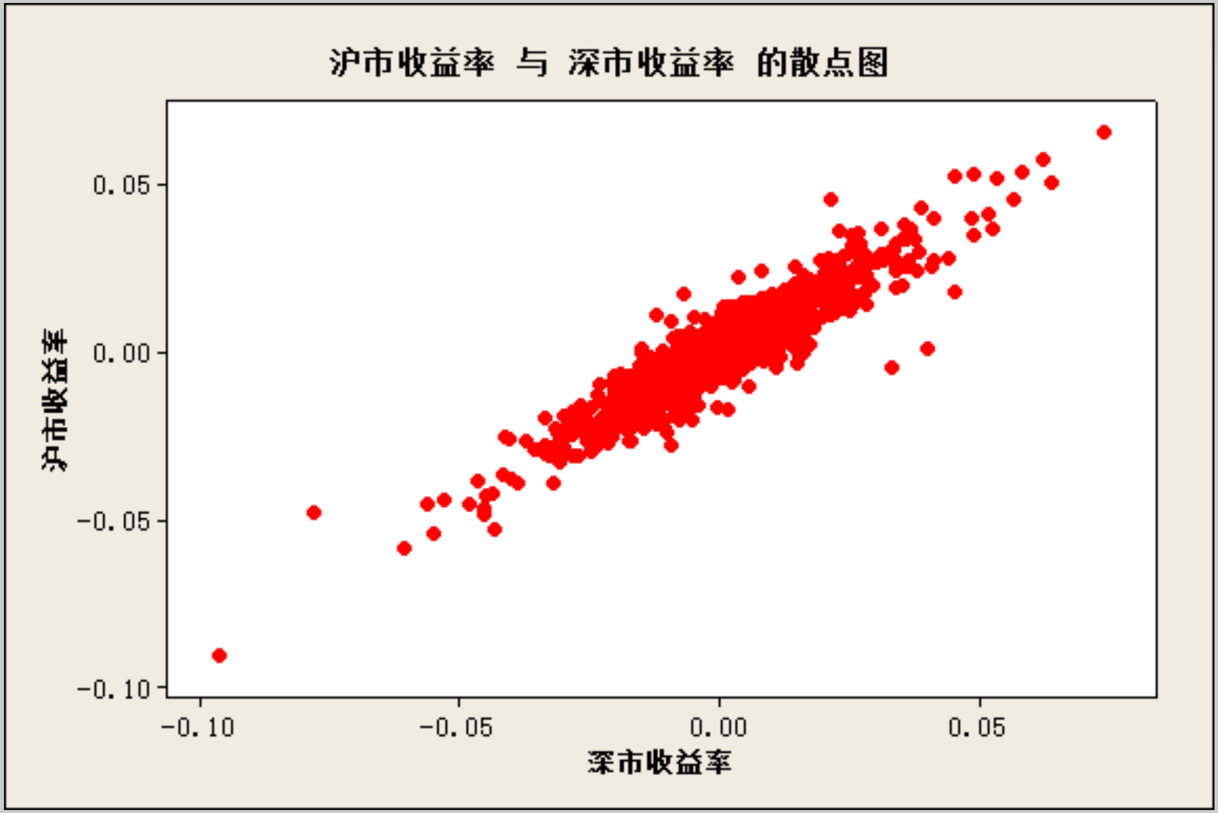

体重 与 月龄 的散点图
体重
月龄
0 10 20 30 40 50 60 70 80
5 10 15 20 25


体重和身高的二元分布
核密度
0.01
0.005
0
体重
身高


体重 与 身高 的散点图
体重
身高
50 60 70 80 90 100 110 120
5 10 15 20 25


二元分布密度等高线图
体重
身高
40 60 80 100 120
0 5 10 15 20 25

【例3.6-2】现有上海和深圳股市同时期日开盘价、最高价、最低价、收盘价、收益率等数据，跨度为2000年1月至2007年4月，各1696组数据。据此绘制沪、深股市收益率的散点图，观察沪、深股市收益率的关系。



二、 随机变量的协方差

1. 协方差的定义

定义1 称 $E\left([X-E(X)][Y-E(Y)]\right)$ 为 X,Y 的协方差. 记为

$$\boxed{\operatorname{cov}(X,Y)=E\left([X-E(X)][Y-E(Y)]\right)}$$

可以证明

$$\boxed{\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)}$$

注：协方差是有量纲的

协方差定义： $\boxed{\operatorname{cov}(X,Y)=E\left([X-E(X)][Y-E(Y)]\right)}$

证明

$$\boxed{\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y).}$$

我们有

$$\begin{aligned}\operatorname{cov}(X,Y)&=E[XY-XE(Y)-YE(X)+E(X)E(Y)]\\&=E(XY)-E(X)E(Y)-E(X)E(Y)+E(X)E(Y)\\&=E(XY)-E(X)E(Y).\end{aligned}$$

2. 协方差的性质

- $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$
- $\operatorname{cov}(X,a)=0$
- $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$
- $\operatorname{cov}(X+Y,Z)=\operatorname{cov}(X,Z)+\operatorname{cov}(Y,Z)$
- $\operatorname{cov}(X,X)=D(X)$
- $D(aX\pm bY)=a^2D(X)+b^2D(Y)\pm 2ab\operatorname{cov}(X,Y)$
- 若X与Y相互独立，则 $\operatorname{cov}(X,Y)=0$

3. X和Y协方差等于0与独立的关系

X和Y独立能推出$\text{cov}(X,Y)=0$,也就是X和Y不相关;

X和Y不相关不能推出X和Y独立;

X和Y独立"直白的说就是X和Y没有任何关系";

X和Y的**相关关系**只是众多关系的一种;

例 1 已知离散型随机向量(X,Y)的概率分布为

X \ Y	−1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

求$\mathrm{cov}(X,Y)$.

例 1 已知离散型随机向量 (X,Y) 的概率分布

X \ Y	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

求 $\mathrm{cov}(X,Y)$.

$E(X)=0\times0.3+1\times0.45+2\times0.25=0.95,$

$E(Y)=(-1)\times0.55+0\times0.25+2\times0.2=-0.15.$

cov(x,y) = E(xy) − E

计算得

$$E(XY)=0\times(-1)\times0.1+0\times0\times0.2+0\times2\times0+1\times(-1)\times0.3+1\times0\times0.5+1\times2\times0.1$$
$$+2\times(-1)\times0.15+2\times0\times0+2\times2\times0.1$$
$$=0.$$

于是 $\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)=0.95\times0.15=0.1425.$

【例3.6-4】 已知二维随机变量(X,Y)的联合概率密度如下，求X与Y的协方差。

$$f(x,y)=\begin{cases}4xy, & 0\le x\le 1,0\le y\le 1\\ 0, & \text{其他}\end{cases}$$

解：
$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy$$

$$=\int_0^1\left(\int_0^1 x\cdot 4xydx\right)dy=2/3=E(Y)$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy=4/9$$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=0$$

三、 随机变量的相关系数

1. 相关系数的定义

$$\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\ \sqrt{D(Y)}}$$

定义2 若$D(X)>0$，$D(Y)>0$，称

$$R(X,Y)=\mathrm{cov}(X^*,Y^*)=\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为X,Y的相关系数。也可记为 $\mathrm{corr}(X,Y)$ 或 ρ_{XY} .

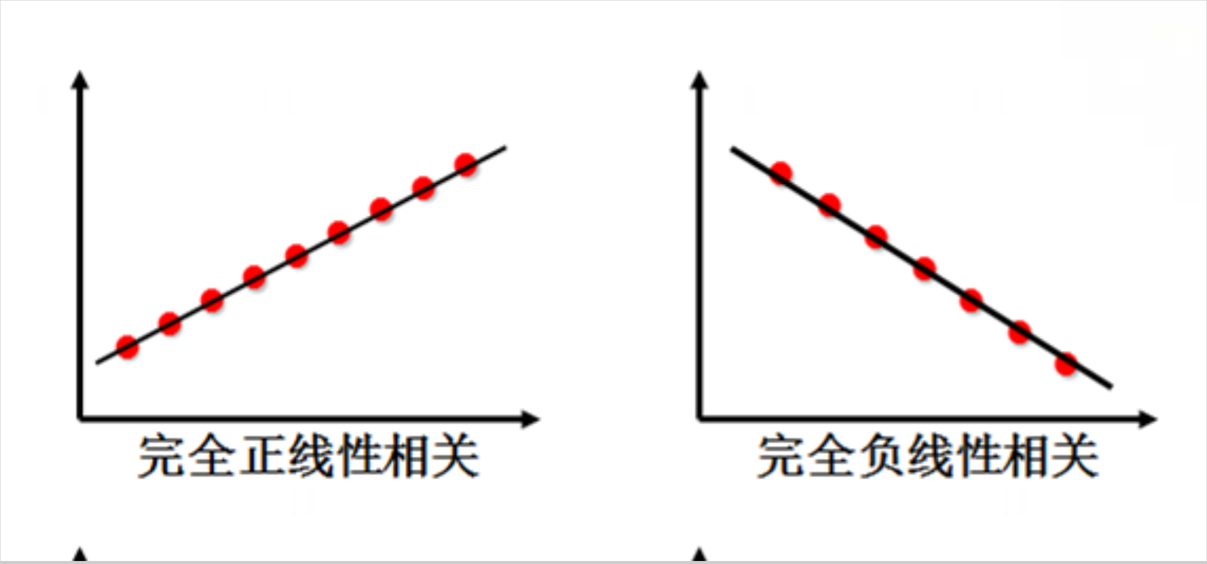
其中

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}},\quad Y^*=\frac{Y-E(Y)}{\sqrt{D(Y)}}$$

注： 相关系数是无量纲的

2. 相关系数的性质

- $|R(X,Y)| \le 1$
- $|R(X,Y)| = 1 \Leftrightarrow$ $Y = a + bX$，且 $R(X,Y) = \begin{cases} 1, & b > 0 \\ -1, & b < 0 \end{cases}$



四、 相关与不相关的定义

1. 定义

这里的相关是指线性相关

若 $\operatorname{cov}(X, Y) = 0$，称 X 与 Y 不相关；

若 $\operatorname{cov}(X, Y) > 0$，称 X 与 Y 正相关；

若 $\operatorname{cov}(X, Y) < 0$，称 X 与 Y 负相关。

2. 不相关与独立之间的关系

$\boxed{X与Y相互独立}$ $\underset{\not\Longleftarrow}{\Longrightarrow}$ $\boxed{X与Y不相关}$

【例3.6-5】 已知随机变量$X\sim U(-1,1)$，$Y=X^2$，求X与Y的相关系数。

解：由于X服从$(-1,1)$上均匀分布，可知 [-1,1]

$$E(X)=0$$

Y=X²

E(X) = 0/2 = 0

$$E(XY)=E(X^3)=\int_{-1}^{1}\frac{x^3}{2}dx=0$$

E(XY) = E(X³) = 0

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=0$$

E(XY) = E(X³)

$$f(x)=\begin{cases}\frac{1}{2}, & -1<x<1\\ 0, & 其它\end{cases}$$

所以

$$R(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=0$$