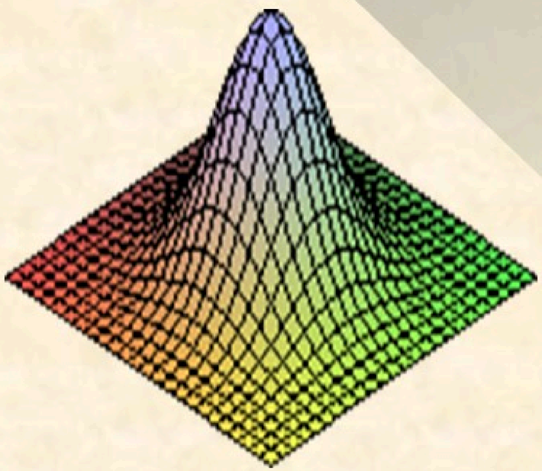
概率论与数理统计



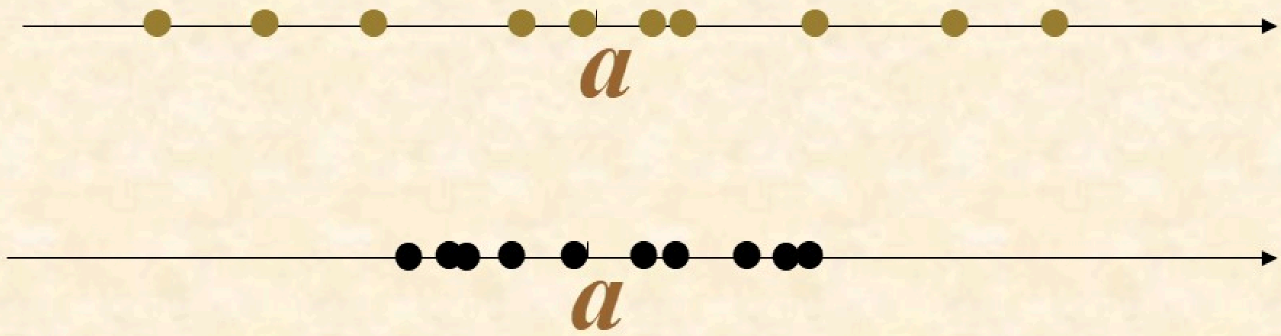
制作人：叶鹰　吴娟

主讲人：吴　娟

wujuan@hust.edu.cn

§ 4.2 方差

例 某零件的真实长度为a，现用甲、乙两台仪器各测量10次，其结果如下：



如何评价两台仪器的优劣？

测量结果的均值相同，都是 a.

但是，乙仪器的测量结果更集中在均值附近. 更优

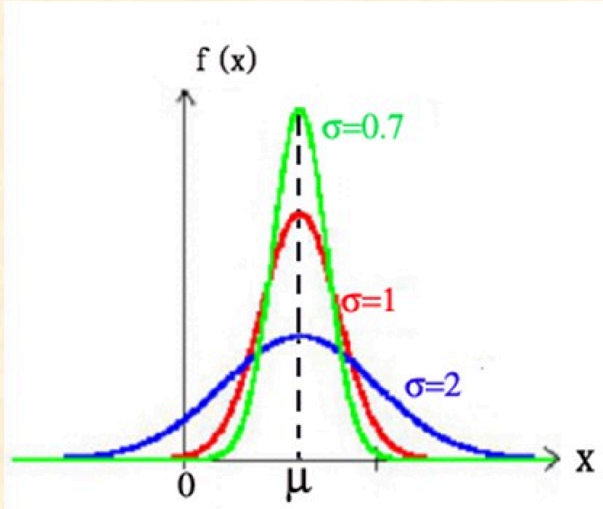
定义　若$E(X^2)$存在，则称$E(X-EX)^2$为X的方差.

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

若X的取值比较集中，则方差较小；

若X的取值比较分散，则方差较大.

若方差$D(X)=0$，则X以概率1取常数值.



定义 若$E(X^2)$存在，则称$E(X-EX)^2$为X的方差.

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

对D.R.V：

$$D(X)=\sum_i [x_i-E(X)]^2 p_i=\sum_i x_i^2 p_i-(EX)^2$$

对C.R.V：

$$D(X)=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)dx=\int_{-\infty}^{+\infty}x^2 f(x)dx-(EX)^2$$

常用分布 数学期望及方差

1. $X \sim B(1, p)$, $E(X) = p$, $DX = p(1-p)$.

2. $X \sim B(n, p)$, $E(X) = np$, $DX = np(1-p)$.

3. $X \sim P(\lambda)$, $EX = \lambda$, $DX = \lambda$.

4. $X \sim U(a,b)$, $E(X) = \dfrac{a+b}{2}$, $DX = \dfrac{(b-a)^2}{12}$.

5. $X \sim E(\lambda)$, $E(X) = 1/\lambda$, $DX = 1/\lambda^2$.

6. $X \sim N(\mu, \sigma^2)$, $EX = \mu$, $DX = \sigma^2$.

方差的基本性质

1. $D(X)\ge 0$ 且 $D(X)=0 \Leftrightarrow P(X=c)=1$ （退化分布）

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

3. $D(X_1+X_2) \xlongequal{X_1与X_2独立} D(X_1)+D(X_2)$

$$=E[X_1+X_2-E(X_1+X_2)]^2$$

$$=E[(X_1-EX_1)^2+2(X_1-EX_1)(X_2-EX_2)+(X_2-EX_2)^2]$$

$$=D(X_1)+D(X_2)+2E[(X_1-EX_1)(X_2-EX_2)]$$

当X_1与X_2独立时， $E[(X_1-EX_1)(X_2-EX_2)]$

$$=E(X_1-EX_1)E(X_2-EX_2)=\mathbf{0}$$

例　设$X \sim B(n, p)$，求$D(X)$.

解　设$X_i \sim B(1, p)$，$i=1,2,\ldots,n$，**相互独立**，则

$$X=\sum_{i=1}^{n} X_i \sim B(n, p)$$

故 $D(X)=\sum_{i=1}^{n} p(1-p)=np(1-p)$

例　设X有期望和方差存在，求EY和DY.

$$\boxed{Y=\frac{X-EX}{\sqrt{DX}} \quad \sim \text{标准化}}$$

解

$$E(Y)=E\left(\frac{X-EX}{\sqrt{DX}}\right)=\frac{1}{\sqrt{DX}}E(X-EX)=0$$

$$D(Y)=D\left(\frac{X-EX}{\sqrt{DX}}\right)=\frac{1}{(DX)}D(X-EX)=1$$

例 设 $X_1, X_2, \cdots, X_n$ 相互独立，$EX_i = \mu, DX_i = \sigma^2$，

$(i = 1, 2, \cdots, n)$，令 $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$，求 $E\bar{X}, D\bar{X}$.

解

$$E\bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu,$$

$$D\bar{X} = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{\sigma^2}{n}.$$

§4.3 随机变量的矩

若 EX^k 存在，称之为X的 k 阶原点矩.

若 $E(X-EX)^k$ 存在，称之为X的 k 阶中心矩.

若 $E(X^kY^l)$存在，称之为X和Y的 $k+l$ 阶混合原点矩.

若$E((X-EX)^k(Y-EY)^l)$存在，称之为X和Y的 $k+l$ 阶混合中心矩.

EX是一阶原点矩，DX是二阶中心矩，

协方差$Cov(X,Y)$是二阶混合中心矩.

§ 4.4 协方差和相关系数

如何简洁描述随机变量之间的关系？

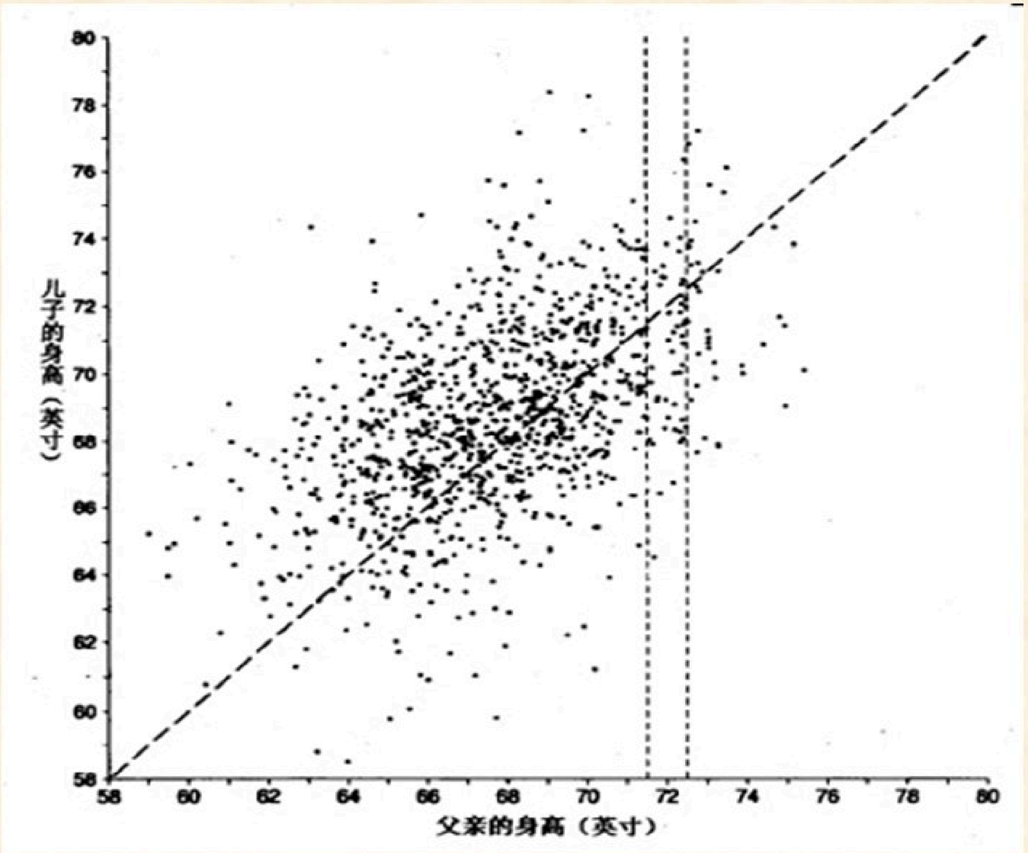
若随机变量X和Y相互独立，则

$$E((X-EX)(Y-EY))=0.$$

若$E((X-EX)(Y-EY))\neq 0$，则

随机变量X和Y不相互独立，而是存在一定的关系.

英国统计学家高尔顿和皮尔逊收集了**1078**个父亲及其成年儿子身高的数据，画出了一张散点图.



问题：父亲及其成年儿子身高是一种什么关系呢？

父亲及其成年男子身高有关系，但没有明确的函数关系.

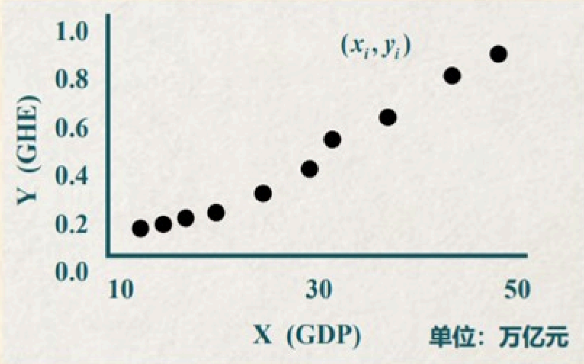
类似的问题有：

吸烟和患肺癌有什么关系？

受教育程度和失业有什么关系？

观察*X*和*Y*变化的一致性

若*X*和*Y*存在"同向"变化关系
——正相关



如何度量*X*和*Y*之间的关系？

选取2003年~2012年国内生产总值*X*(GDP)、政府卫生支出*Y*(GHE)数据记录如下表所示:

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
X(GDP)	13	16	18	22	27	32	34	40	47	52
Y(GHE)	0.11	0.13	0.16	0.18	0.26	0.36	0.48	0.57	0.75	0.84

度量各随机点到平均位置的偏差

- 平均位置

$$(\bar{x},\bar{y})=(10,0.38)$$

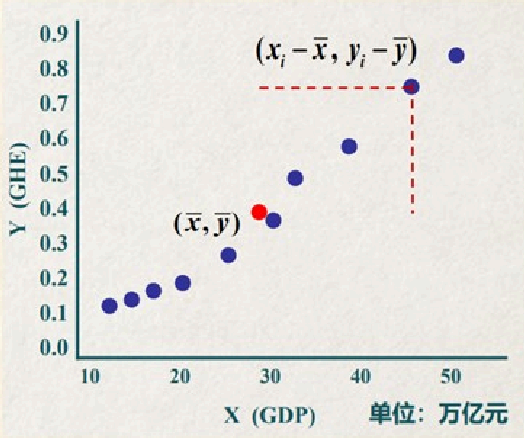
- 点(x_i,y_i)到$(\bar{x},\bar{y})$的偏差

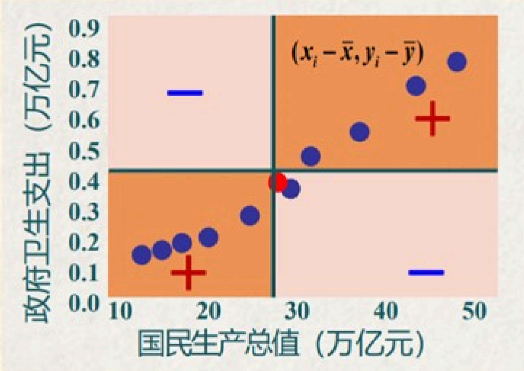
$$(x_i-\bar{x})(y_i-\bar{y})$$

- 平均偏差

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

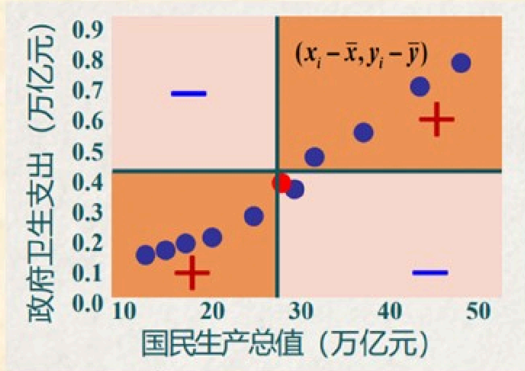
$$=3>0$$





如何度量X和Y之间的关系？

平均偏差 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$



协方差的定义

$$Cov(X,Y)=E\big((X-EX)(Y-EY)\big)$$

➢ 若X与Y相互独立，则 $Cov(X,Y)=0$.

协方差的符号反映X和Y变化的一致性.

一、协方差

$$Cov(X,Y)=E[(X-EX)(Y-EY)]$$
$$=E(XY)-(EX)(EY)$$

基本性质

1° $Cov(X,Y)=Cov(Y,X)$

2° $Cov(aX,bY)=abCov(X,Y)$

3° $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

***X*与*Y*不相关：*Cov*(*X*, *Y*)=0**

定理1 下面等式等价

（1）$Cov(X, Y)=0$；

（2）$E(XY)=E(X)E(Y)$；

（3）$D(X+Y)=D(X)+D(Y)$.

证　由 $Cov(X,Y)=E(XY)-(EX)(EY)$

和 $D(X+Y)=D(X)+D(Y)+2\,Cov(X,Y)$ 即得

定理2 X与Y相互独立且协方差存在，则X与Y不相关，但反之不然.

证　由期望、方差的性质及定理1即得

例 设$X \sim N(0,1)$，$Y=X^2$，求$Cov(X,Y)$.

解

$$Cov(X,Y)=E(XY)-E(X)\cdot E(Y)$$

$$=E(X^3)-0\times E(Y)=\int_{-\infty}^{+\infty}\frac{x^3}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0$$

相关系数

$$Cov(X,Y) = E[(X-EX)(Y-EY)]$$ 与量纲有关

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$ 与量纲无关

$$\rho_{XY} = Cov\left(\frac{X-EX}{\sqrt{DX}}, \frac{Y-EY}{\sqrt{DY}}\right) = \frac{Cov(X,Y)}{\sqrt{Cov(X,X)}\sqrt{Cov(Y,Y)}}$$