引例

头像识别技术有着重要应用,其原理是什么呢?

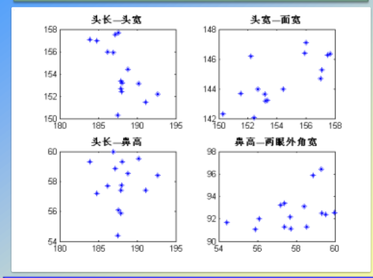
为了制定中国成年人(男性)头型系列标准，在15个省市取样，每个人的头部共测了42项指标。

现选部分指标进行分析。

地区	X1(头长)	X2(头宽)	X5(面宽)	X8(鼻高)	X11(两眼外角宽)
黑龙江	184.8	157	144.7	57.2	93.2
吉林	183.91	157.09	145.28	59.32	92.51
北京	187.16	157.51	146.28	58.87	95.86
陕西	192.7	152.2	146.2	58.4	93.1
江苏	186.94	155.93	146.38	59.97	92.55
山东	186.2	156	147.1	57.7	92.2
河北	188.1	153.2	143.2	59.3	96.4
湖南	187.93	153.36	143.24	55.89	91.09
湖北	187.59	157.67	146.37	56.09	92.04
广东	187.97	152.42	142.06	57.73	91.12
广西	187.5	150.3	142.3	54.4	91.7
福建	187.9	152.7	144	57.4	93.4
四川	188.85	154.43	143.99	58.55	91.27
贵州	191.1	151.5	143.7	57.4	91.3
云南	190.22	153.16	143.66	59.54	92.39

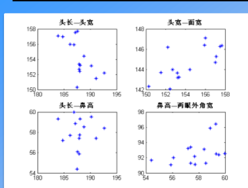
分析变量间关系

中国成年男性头型部分指标



给出你的猜想

如何描述变量间关系?



通过观察可以发现:
有些变量间没有明显关系，例如头长与鼻高，其图像没有明显规律。

有些变量间存在关系:
如头长与头宽，有反向增长关系;
而头宽与面宽，有同向增长关系。

考虑变量的乘积? 离差的乘积?

$|\rho|\le 1$的证明

1) $|\rho|\le 1$

证明: $0\le D(X^*\pm Y^*)=D(X^*)+D(Y^*)\pm 2cov(X^*,Y^*)$

$=2\pm 2\rho_{XY} \quad =2(1\pm 1\rho_{XY})$

$1\pm 1\rho_{XY}\ge 0$

$\therefore |\rho_{XY}|\le 1$

$|\rho|=1$的充要条件

2) $|\rho|=1 \Longleftrightarrow X$与$Y$依概率为1线性相关，即

$\exists \alpha,\beta\ (\alpha\ne 0)\quad s.t$

$P\{Y=\alpha X+\beta\}=1$

证明: "⇒" 必要性

$\rho=-1$时

$D(X^*+Y^*)=D(X^*)+D(Y^*)+2\text{cov}(X^*,Y^*)$

$=2+2\rho_{X^*Y^*}=0$

$E(X^*+Y^*)=0$

$|\rho|=1$的充要条件

2) $|\rho|=1 \Longleftrightarrow X$与$Y$依概率为1线性相关，即

$\exists \alpha,\beta\ (\alpha\ne 0)\quad s.t$

$P\{Y=\alpha X+\beta\}=1$

证明: "⇒" 必要性

由方差的性质4) 得

$P\{X^*+Y^*=E(X^*+Y^*)\}=1$ 即

$P\{X^*+Y^*=0\}=1$

$$P\left\{Y=-\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}X+\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}E(X)+E(Y)\right\}=1$$

对$\rho=1$ 同理可得。

"⇐" 充分性

$$P\{Y=\alpha X+\beta\}=1$$
$$E(Y)=\alpha E(X)+\beta$$
$$D(Y)=\alpha^2 D(X)$$
$$cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$
$$=E\{[X-E(X)][\alpha X+\beta-E(\alpha X+\beta)]\}$$
$$=\alpha D(X)$$
$$\rho_{XY}=\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\alpha}{\sqrt{\alpha^2}}=\pm 1$$

线性变换对相关系数的影响

3）若$\xi=a_1X+b_1$，$\eta=a_2Y+b_2$ 则

$$\rho_{\xi\eta}=\frac{a_1a_2}{|a_1a_2|}\rho_{XY}$$

证明：$D(\xi)=a_1^2D(X)\quad D(\eta)=a_2^2D(Y)$

$$\mathrm{cov}(\xi,\eta)=E\{[\xi-E(\xi)][\eta-E(\eta)]\}$$
$$=E\{[a_1X-a_1E(X)][a_2Y-a_2E(Y)]\}$$
$$=a_1a_2E\{[X-E(X)][Y-E(Y)]\}$$
$$=a_1a_2\,\mathrm{cov}(X,Y)$$
$$\rho_{\xi\eta}=\frac{\mathrm{cov}(\xi,\eta)}{\sqrt{D(\xi)}\sqrt{D(\eta)}}=\frac{a_1a_2}{\sqrt{(a_1a_2)^2}}\rho_{XY}=\frac{a_1a_2}{|a_1a_2|}\rho_{XY}$$

不相关但也不独立的例子

例：(X,Y)在以原点为圆心的单位圆内服从均匀分布。

$$f(x,y)=\begin{cases}\frac{1}{\pi} & x^2+y^2\le 1\\ 0 & \text{其它}\end{cases}$$

$$f_X(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2\sqrt{1-x^2}}{\pi}\qquad -1\le x\le 1$$
$$f_Y(y)=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{\pi}dx=\frac{2\sqrt{1-y^2}}{\pi}\qquad -1\le y\le 1$$
$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx=\int_{-1}^{1}\frac{2x\sqrt{1-x^2}}{\pi}dx=0$$

同理：$E(Y)=0$

协方差.相关系数与矩

例：(X,Y)在以原点为圆心的单位圆内服从均匀分布。

$$f(x,y)=\begin{cases}\frac{1}{\pi} & x^2+y^2\le 1\\ 0 & \text{其它}\end{cases}$$

$$E\{[X-E(X)][Y-E(Y)]\}=\iint_{x^2+y^2\le 1}\frac{xy}{\pi}dxdy$$
$$=\frac{1}{\pi}\int_0^1\int_0^{2\pi}r^3\sin\theta\cos\theta\, d\theta dr=0$$

可以验证 $D(X)>0,\ D(Y)>0$

从而 $\rho_{XY}=0$

但 $f(x,y)\ne f_X(x)f_Y(y)$

不相关 不一定 相互独立

例4.3.2

假二维随机变量(X,Y)在矩形 $G=\{(x,y)|0\le x\le 2,\ 0\le y\le 1\}$上服从均匀分布.

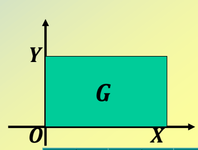
记 $U=\begin{cases}0 & X\le Y\\ 1 & X>Y\end{cases}\qquad V=\begin{cases}0 & X\le 2Y\\ 1 & X>2Y\end{cases}$

求ρ_{UV}

分析：$\rho_{UV}=\dfrac{cov(U,V)}{\sqrt{D(U)}\sqrt{D(V)}}=\dfrac{E(UV)-E(U)E(V)}{\sqrt{D(U)}\sqrt{D(V)}}$

关键是求$E(UV)$

⟶ 求出UV分布律



例4.3.2

假二维随机变量(X,Y)在矩形 $G=\{(x,y)|0\le x\le 2,\ 0\le y\le 1\}$上服从均匀分布.

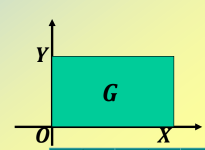
记 $U=\begin{cases}0 & X\le Y\\ 1 & X>Y\end{cases}\qquad V=\begin{cases}0 & X\le 2Y\\ 1 & X>2Y\end{cases}$

求ρ_{UV}

解：由已知可得 $f(x,y)=\begin{cases}1/2 & (x,y)\in G\\ 0 & \text{其它}\end{cases}$

$$E(U)=0\times P\{X\le Y\}+1\times P\{X>Y\}$$
$$=\int_0^1\left[\int_y^2 f(x,y)dx\right]dy=3/4$$
$$D(U)=E(U^2)-[E(U)]^2=3/16$$

同理 $E(V)=1/2\quad D(V)=1/4$



UV 的分布律为：

$$UV=\begin{cases}0 & X\le 2Y\\ 1 & X>2Y\end{cases}=V$$

故 $E(UV)=E(V)=1/2$

从而 $\rho_{UV}=\dfrac{\mathrm{cov}(U,V)}{\sqrt{D(U)}\sqrt{D(V)}}=\dfrac{E(UV)-E(U)E(V)}{\sqrt{D(U)}\sqrt{D(V)}}$

$$=\frac{\frac{1}{2}-\frac{3}{4}\times\frac{1}{2}}{\sqrt{\frac{3}{16}\times\frac{1}{4}}}=\frac{\sqrt{3}}{3}$$

例4.3.3

某集装箱中放有100件产品，其中一、二、三等品分别为80、10、10件。现从中任取一件，记

$$X_i=\begin{cases}1 & \text{抽到}i\text{等品}\\ 0 & \text{其它}\end{cases}\quad i=1,2,3\quad \text{求}\rho_{X_1X_2}$$

分析：$\rho_{X_1X_2}=\dfrac{\mathrm{cov}(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=\dfrac{E(X_1X_2)-E(X_1)E(X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$

关键是求$E(X_1X_2)$ $\longrightarrow$ 求出X_1X_2分布律

某集装箱中放有100件产品，其中一、二、三等品分别为80、10、10件。现从中任取一件，记

$$X_i=\begin{cases}1 & \text{抽到}i\text{等品}\\ 0 & \text{其它}\end{cases}\quad i=1,2,3\quad \text{求}\rho_{X_1X_2}$$

解：由已知可得

$E(X_1)=0\times P\{\text{抽到非一等品}\}+1\times P\{\text{抽到一等品}\}=0.8$

$D(X_1)=0.8(1-0.8)=0.16$

同理 $E(X_2)=0.1\quad D(X_2)=0.09$

X_1X_2 的取值为：

$$X_1X_2=\begin{cases}1 & \text{抽到的为一等品且为二等品}\\ 0 & \text{其它}\end{cases}$$

$P\{X_1X_2=1\}=P\{\phi\}=0$

$E(X_1X_2)=1\cdot P\{X_1X_2=1\}+0\cdot P\{X_1X_2=0\}$
$=0$

从而

$$\rho_{X_1X_2}=\frac{\mathrm{cov}(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=\frac{E(X_1X_2)-E(X_1)E(X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$$

$$=-\frac{2}{3}$$