Ch7 積差相關 [相關係數]

1. 成對的資料 $(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)$
欲測量他們的相關性

定義

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

為 X 與 Y 的相關係數

2. $-1 \le r \le 1$

(証) 令 $\vec{a} = \langle x_1-\bar{x}, \cdots\cdots x_n-\bar{x} \rangle$

$\vec{b} = \langle y_1-\bar{y}, \cdots\cdots y_n-\bar{y} \rangle$

定義: $\vec{a}\cdot\vec{b} = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$

$|\vec{a}| = \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})}$ $|\vec{b}| = \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})}$

$$r=\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}=\frac{\vec{a}}{|\vec{a}|}\frac{\vec{b}}{|\vec{b}|}$$ 故 $-1\le r\le 1$

3. $(x_1, y_1)\cdots(x_n, y_n)$ 愈接近直線

則 $|r|$ 愈靠近 1

$(Z_{1i}, Z_{2i})\ (Z_{12}, Z_{22})\cdots(Z_{1n}, Z_{2n})$

4. 資料標準化後，令

$$Z_{1i}=\frac{x_i-\bar{x}}{\sigma_1}\qquad Z_{2i}=\frac{y_i-\bar{y}}{\sigma_2}$$

此處 $\sigma_1=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

$$\sigma_2=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}$$

Z_1, Z_2 的相關係數仍為 r

(f)

$$令\ r' = \frac{\sum_{i=1}^{n} (Z_{1i} - Z_{2i})}{\sqrt{\sum_{i=1}^{n} Z_{1i}^2 \sum_{i=1}^{n} Z_{2i}^2}}$$

$$= \frac{\sum_{i=1}^{n} \frac{(x_1 - \overline{X})}{\sigma_1} \frac{(Y_i - \overline{Y})}{\sigma_2}}{\sqrt{\frac{\sum_{i=1}^{n} (x_1 - \overline{X})}{\sigma_1^2}} \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{\sigma_2^2}}}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{X}) \sum_{i=1}^{n} (Y_i - \overline{Y})}}$$

$$= r$$

5. 当 $Y_i = aX_i + b \quad i = 1, 2, \cdots n$

$$r = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

当是水平线
相关系数
为0

(pf)

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (aX_i + b - a\overline{X} - b)^2$$

$$= a^2 \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$= \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

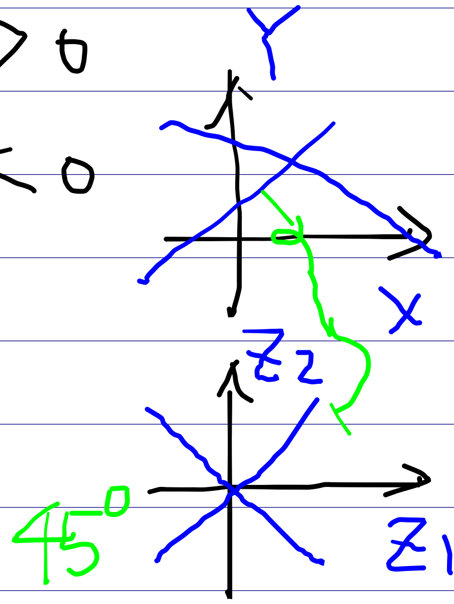
$$= a^2 \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})$$

$$= a \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X}) \sum_{i=1}^{n} (Y_i - \overline{Y})}}$$

$$= \frac{a\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\, a^2\sum_{i=1}^{n}(x_i-\overline{x})^2}}$$

$$= \frac{a}{|a|} = \begin{cases} 1 & a>0 \\ -1 & a<0 \end{cases}$$



6. 當 $Y_i = aX_i + b$

$r=1$
$=-1$ $\Leftarrow Z_{2i} = \begin{cases} Z_{1i}, & a>0 \\ -Z_{1i}, & a<0 \end{cases}$ $i=1,2,\dots n$

$$Z_{2i} = \frac{Y_i-\overline{Y}}{\sigma_2} = \frac{a(x_i-\overline{x})}{|a|\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}} = \frac{a}{|a|}$$

標準化

這只有二種 $= Z_{1i} = \begin{cases} Z_{1i} & a>0 \\ -Z_{1i} & a<0 \end{cases}$

正45°

負45°

/ ⇒ 可靠度

(如果今天加分，晚一天作也應是加分)

⇒ 有效反應出要評量的內容。

難度、鑑別度

△ 校標關聯 r