$\Rightarrow$ この直線は

$$\begin{cases} y-\bar{y}=\hat{a}(x-\bar{x}) \\ \hat{a}=\dfrac{S_{yy}-S_{xx}+\sqrt{(S_{xx}-S_{yy})^2+4S_{xy}^2}}{2S_{xy}} \end{cases}$$

(証明は WWW ページ)

で，$\hat{g}$ の最小値 $\hat{g}_m$ は

$$\hat{g}_m=\frac{1}{2}(S_{xx}+S_{yy})(1-\hat{r})$$

ただし $\hat{r}=\dfrac{\sqrt{(S_{xx}-S_{yy})^2+4S_{xy}^2}}{S_{xx}+S_{yy}}$ $(0\leq\hat{r}\leq 1)$

◎ $\hat{g}_m$，$\hat{r}$ は回転不変性をもつ。

(例えば $S_{xx}+S_{yy}=\sum_{k=1}^{n}\{(x_k-\bar{x})^2+(y_k-\bar{y})^2\}$

$=(x_k, y_k)$ と $(\bar{x},\bar{y})$ の距離の平方和 ← 回転不変)

よって $\dfrac{\hat{g}_m}{S_{xx}+S_{yy}}=\dfrac{1}{2}(1-\hat{r})$ は

- 散布図のスケールにかかわらない。 (x_k, y_k をすべて α 倍しても不変)
- 直線相関が 強い $\Leftrightarrow$ 小さい
- 〃 弱い $\Leftrightarrow$ 大きい

$\Rightarrow$ $\hat{r}$ を "相関係数" として使えるのでは？

(注 r も g の最小値 $g_m=\dfrac{S_{xx}S_{yy}-S_{xy}^2}{S_{xx}}$ の対称性を考慮してスケールを消したもの

$\left(\dfrac{g_m}{S_{yy}}=\dfrac{S_{xx}S_{yy}-S_{xy}^2}{S_{xx}S_{yy}}=1-r^2\right.$ から得られる

$\hat{r}=1 \Leftrightarrow (S_{xx}-S_{yy})^2+4S_{xy}^2=(S_{xx}+S_{yy})^2$

$\Leftrightarrow S_{xy}^2=S_{xx}S_{yy} \Leftrightarrow |r|=1 \Leftrightarrow$ 完全な直線相関

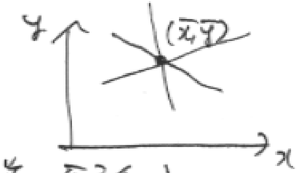
$\hat{r}=0 \Leftrightarrow (S_{xx}-S_{yy})^2+4S_{xy}^2=0$

$\Leftrightarrow S_{xx}=S_{yy}$ かつ $S_{xy}=0$

$\Leftrightarrow$ $\hat{g}(a, \underbrace{\bar{y}-a\bar{x}}_{\Downarrow})$ の値は a によらない (証明は WWW ページ)

($b=\bar{y}-a\bar{x}$ は b に関する最小値を与え，これは 直線が $y-\bar{y}=a(x-\bar{x})$ の形で あることを意味する.

$\Leftrightarrow$ $(\bar{x},\bar{y})$ を通るどの方向の直線に対しても $\hat{g}$(= 直線とデータの距離の平方和) が 変化せず一定 となる



$\Leftrightarrow$ どの方向にも直線相関がない

∴ r のかわりに $\hat{r}$，a のかわりに $\hat{a}$ を使えば 疑問 1o, 2o, 3o は 解決する.

しかし，$\hat{r}$ は x, y が「同種でない」データだと弱く，x, y 別々のスケール変換に対して不変でない