マハラノビス距離

$(\boldsymbol{x_k}, \boldsymbol{y_k})$と$(\boldsymbol{x_n}, \boldsymbol{y_n})$の２点間距離を算出

$$D(k,n) = \sqrt{\begin{bmatrix}\boldsymbol{x_k - x_n} & \boldsymbol{y_k - y_n}\end{bmatrix}\begin{bmatrix}\boldsymbol{V_x} & \boldsymbol{S_{xy}} \\ \boldsymbol{S_{xy}} & \boldsymbol{V_y}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{x_k - x_n} \\ \boldsymbol{y_k - y_n}\end{bmatrix}}$$

基準化すると、平均値0、標準偏差1であるから、$\boldsymbol{V_x} = \boldsymbol{V_y} = 1$

$$相関係数\boldsymbol{r} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$= \frac{\boldsymbol{S_{xy}}}{\sqrt{\boldsymbol{V_x}} \times \sqrt{\boldsymbol{V_y}}} = \boldsymbol{S_{xy}}$$

$$D(k,n) = \sqrt{\begin{bmatrix}\boldsymbol{x_k - x_n} & \boldsymbol{y_k - y_n}\end{bmatrix}\begin{bmatrix}\boldsymbol{1} & \boldsymbol{r} \\ \boldsymbol{r} & \boldsymbol{1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{x_k - x_n} \\ \boldsymbol{y_k - y_n}\end{bmatrix}}$$

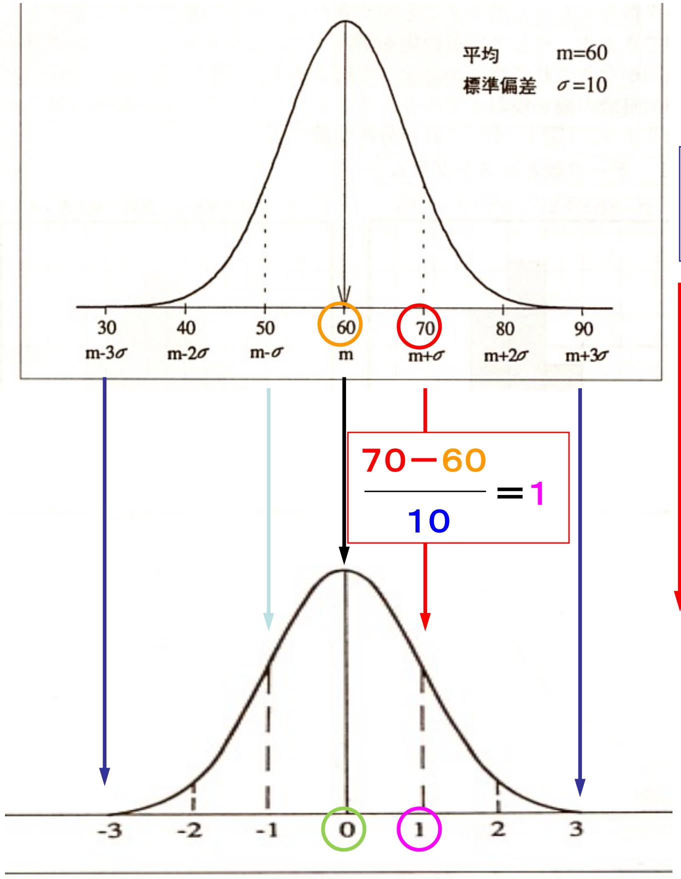
２変量のとき

$$D=\sqrt{\begin{bmatrix}x-\bar{x} & y-\bar{y}\end{bmatrix}\begin{bmatrix}V_x & S_{xy}\\ S_{xy} & V_y\end{bmatrix}^{-1}\begin{bmatrix}x-\bar{x}\\ y-\bar{y}\end{bmatrix}}$$

１変量のとき

$$D=\sqrt{{}^{T}xV^{-1}x}=\sqrt{(x-\bar{x})(\sigma_x{}^2)^{-1}(x-\bar{x})}=\sqrt{\frac{(x-\bar{x})^2}{\sigma_x{}^2}}=\boxed{\frac{|x-\bar{x}|}{\sigma_x}}$$

平均体重が60kg、標準偏差が10kgの分布



$$\frac{70-60}{10}=1$$

平均 m＝60

標準偏差σ＝10

基準化（標準化）★

$$統計量T=\frac{データ-平均値}{標準偏差}$$

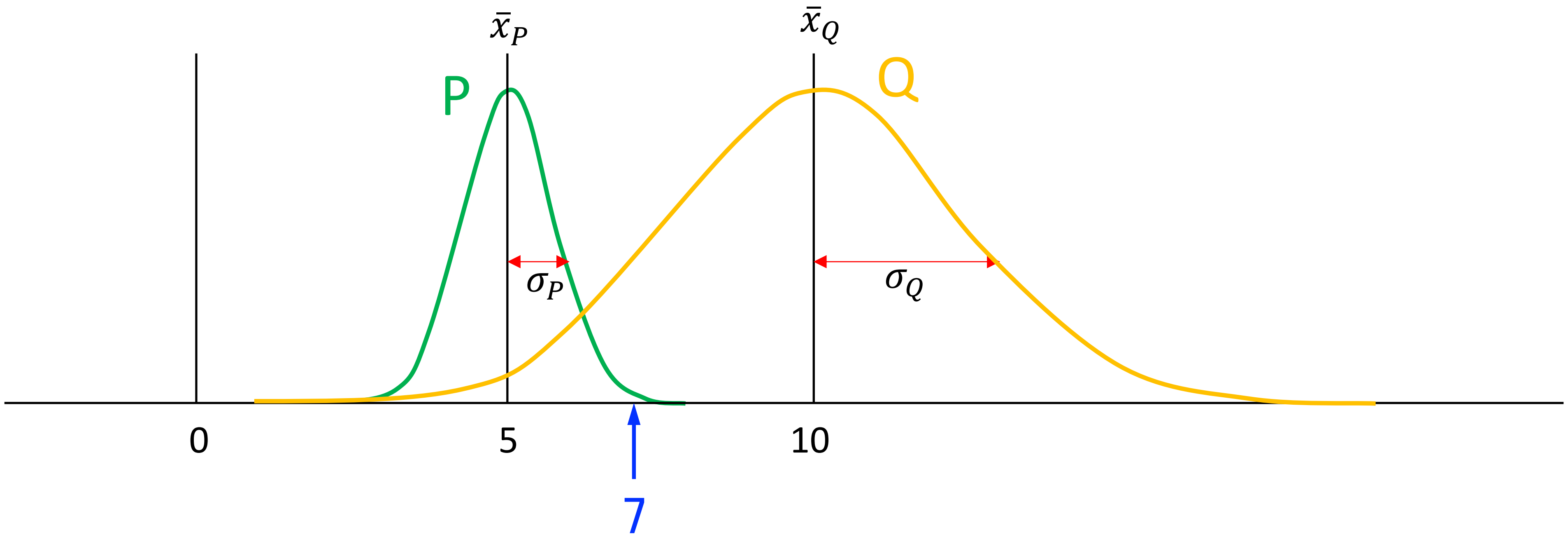
平均 m＝0

標準偏差σ＝1

グループP： 平均値$\bar{x}_P = 5$、標準偏差$\sigma_P = 1$

グループQ： 平均値$\bar{x}_Q = 10$、標準偏差$\sigma_Q = 3$ のとき

$x = 7$の人はどちらのグループ？



$$D_P = \frac{x - \bar{x}_P}{\sigma_P} = \frac{|7 - 5|}{1} = 2$$

$$D_Q = \frac{x - \bar{x}_Q}{\sigma_Q} = \frac{|7 - 10|}{3} = 1$$

以上より、$x = 7$はグループQに近い

	$\bar{x}$	$\bar{y}$	${\sigma_x}^2$	${\sigma_y}^2$	σ_{xy}
P	3	4	9	7	6
Q	8	11	7	6	8

$$D = \sqrt{\begin{bmatrix} x-\bar{x} & y-\bar{y} \end{bmatrix} \begin{bmatrix} V_x & S_{xy} \\ S_{xy} & V_y \end{bmatrix}^{-1} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}}$$

$x=5$、$y=7$の人はどちらのグループ？

$$D_P = \sqrt{\begin{bmatrix} 5-3 & 7-4 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 6 & 7 \end{bmatrix}^{-1} \begin{bmatrix} 5-3 \\ 7-4 \end{bmatrix}}$$

$$= \sqrt{(2\ \ 3)\frac{1}{9\times7-6^2}\begin{pmatrix} 7 & -6 \\ -6 & 9 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix}}$$

$$= \sqrt{\frac{1}{27}(2\ \ 3)\begin{pmatrix} 2 \\ 15 \end{pmatrix}} = \sqrt{\frac{49}{27}} = 1.35$$

$$D_Q = \sqrt{\begin{bmatrix} 5-8 & 7-11 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 8 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 5-8 \\ 7-11 \end{bmatrix}}$$

$$= \sqrt{(-3\ \ -4)\frac{1}{7\times6-8^2}\begin{pmatrix} 6 & -8 \\ -8 & 7 \end{pmatrix}\begin{pmatrix} -3 \\ -4 \end{pmatrix}}$$

$$= \sqrt{\frac{-1}{22}(-3\ \ -4)\begin{pmatrix} 14 \\ -4 \end{pmatrix}} = \sqrt{\frac{26}{22}} = 1.09$$

$D_P > D_Q$なので、Qグループに近い