- **モード（最頻値）：**

  最も多い度数の階級値 $\longrightarrow$ 表 1.3 (P.7) のデータでは 5.45，階級の幅によって変わる

2.4　相関係数 (P.23)

2 変数データの組に関する代表値 $\Longrightarrow$ 共分散，相関係数

例： 100 人の家計からの消費と所得，身長と体重

n 組のデータ (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n)

共分散 s_{xy}

$$s_{xy} = \frac{1}{n}\Big((x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y})$$

$$
\begin{aligned}
&\quad + \cdots + (x_n - \overline{x})(y_n - \overline{y})\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}
\end{aligned}
$$

$s_{xy} > 0$： **正の相関 (x と y との関係はプラスの傾き)**

$s_{xy} < 0$： **負の相関 (x と y との関係はマイナスの傾き)**

$s_{xy} = 0$： **相関なし (x と y との関係は正負の傾きを決定できず)**

相関 $\Longrightarrow$ 互いにかかわりを持つこと。相互に関係しあっていること。(『国語大辞典 (新装版)』小学館，1988)

相関の強弱を表す指標 $\Longrightarrow$ 相関係数 r

$$r = \frac{s_{xy}}{s_x s_y}$$

ただし，

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2,$$

とし，s_x, s_y は x の標準偏差，y の標準偏差である。

$r > 0$： **正の相関 (x と y との関係はプラスの傾き)**

$r < 0$： **負の相関 (x と y との関係はマイナスの傾き)**

$r = 0$： **相関なし** (x **と** y **との関係は正負の傾きを決定できず**)

r **は，**

$$-1 \leq r \leq 1$$

となる。

証明：

次のような t **に関する式を考える。**

$$f(t) = \frac{1}{n}\sum_{i=1}^{n}\Big((x_i - \overline{x})t - (y_i - \overline{y})\Big)^2,$$

平方和なので，必ずゼロ以上となる。よって，すべての t **について，** $f(t) \geq 0$ **となるための**

条件を求めればよい。t に関する２次方程式の判別式がゼロ以下となる条件を求める。

$$
\begin{aligned}
f(t) &= t^2 \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\
&\quad - 2t \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 \\
&= s_x^2 t^2 - 2 s_{xy} t + s_y^2 \geq 0
\end{aligned}
$$

判別式

$$
\frac{D}{4} = s_{xy}^2 - s_x^2 s_y^2 \leq 0
$$

$$
\frac{s_{xy}^2}{s_x^2 s_y^2} \leq 1,
$$

$$-1 \leq \frac{s_{xy}}{s_x s_y} \leq 1,$$

を得る。

r が 1 に近いほど，正の相関が強くなる (x と y のプロットが正の傾きで一直線上に近づく)。

r が -1 に近いほど，負の相関が強くなる (x と y のプロットが負の傾きで一直線上に近づく)。

$r = -1, 1$ のとき，x と y は一直線上に並ぶ ($r = 1$ は正の傾き，$r = -1$ は負の傾き)。