



# 第6章 相関分析

## § 1 相関関係の把握

- 1 散布図による視覚的把握
- 2 相関係数による数值的把握

## § 2 母相関係数の検定

- 1 母相関係数の検定
- 2 母相関係数の推定

## § 3 相関係数のジャックナイフ推定



# § 1 相関関係の把握

- 1 散布図による視覚的把握
- 2 相関係数による数値的把握

# 1 散布図による視覚的把握

【例題6-1】ある製品の強度  $y$  と硬化剤の量  $x$  の関係。

## ■ 考え方と適用手法

2つの量的な変数 ( $x$ 、 $y$ )  $\Rightarrow$  散布図

## ■ 相関関係

相関関係:  $x$  の変化に伴って、 $y$  も変化する関係

正の相関:  $x$  が増えると  $y$  も増えるような関係

負の相関:  $x$  が増えると  $y$  は減るような関係

相関関係  $\neq$  因果関係

相関関係 データにもとづいて検討できる

因果関係 データの背後の学理的知識や技術的知識から

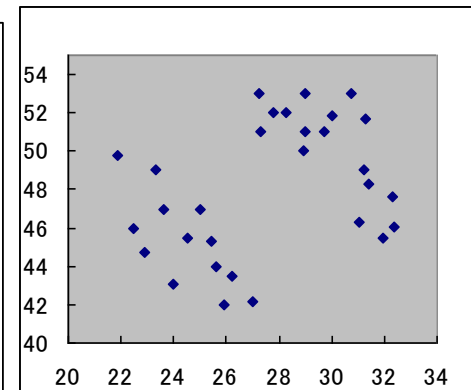
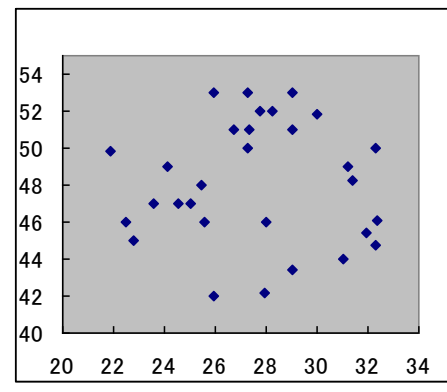
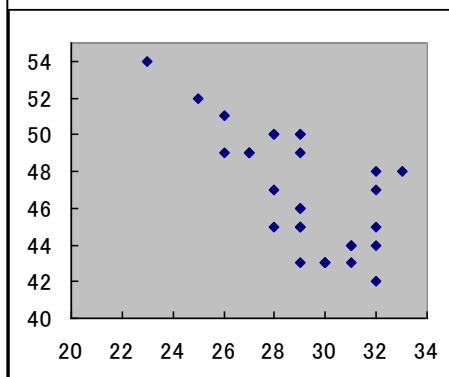
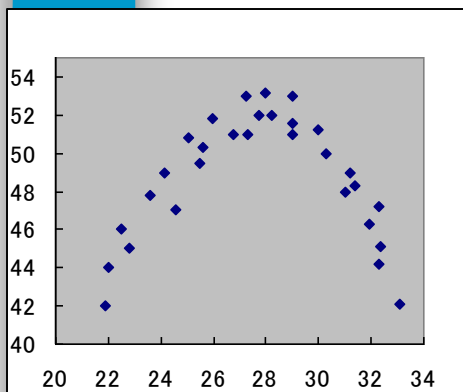
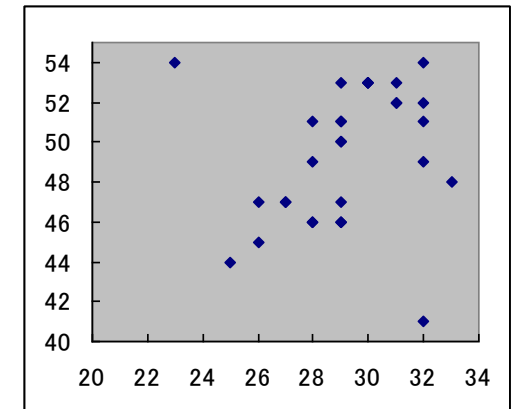
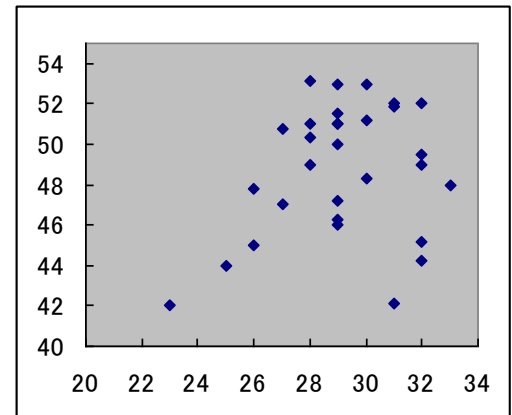
## ■ 散布図

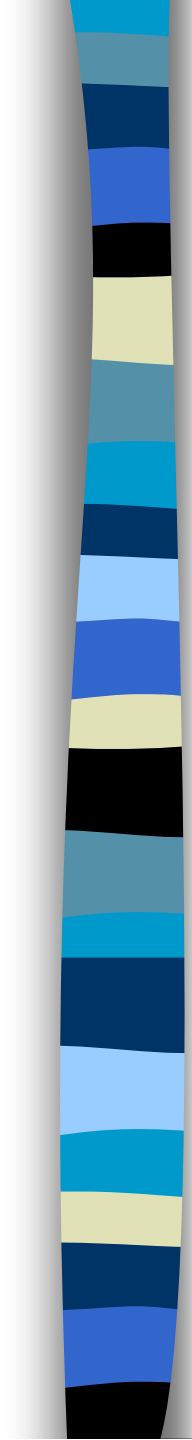
2つの変数の一方を横軸に、もう一方を縦軸にプロット

# 散布図の作成

## 散布図視察のポイント

- ①はずれ値はないか
- ②2つの変数にはどのような関係  
直線、曲線、正負、無相関
- ③グループは形成されていないか





【例題6-2】例題6-1のデータに製品の種類(A・B)の情報追加

## ■ 考え方と適用手法

①種類別に表示

②種類ごとに異なるマークで ← 層別散布図

## ■ 層別散布図の作成

EXCEL

## 2 相関係数による数値的把握

【例題6-3】例題6-1の強度と硬化剤の量の相関係数

### ■ 考え方と適用手法

2変数の間の相関関係の数値判断 ⇒ 相関係数  $r$

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

$r > 0$  正の相関

$r < 0$  負の相関

$$r = 0$$

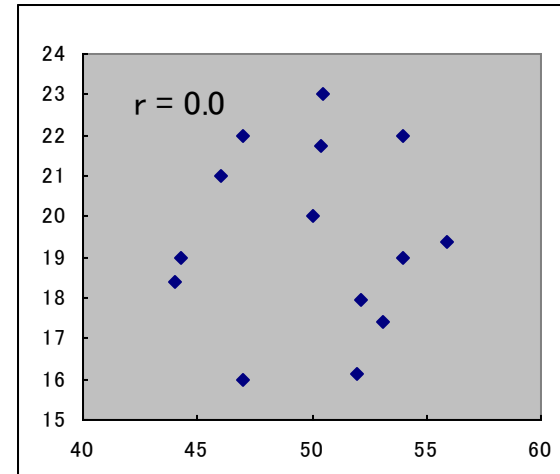
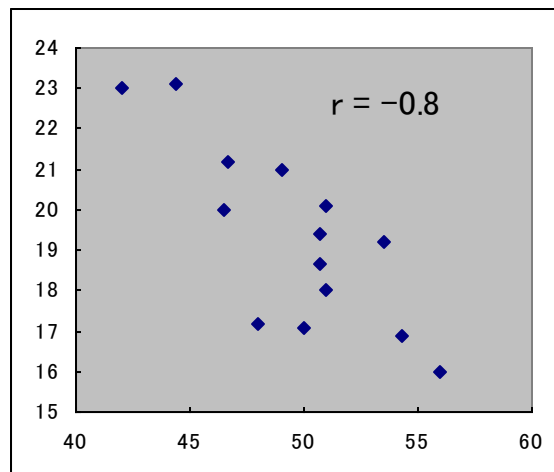
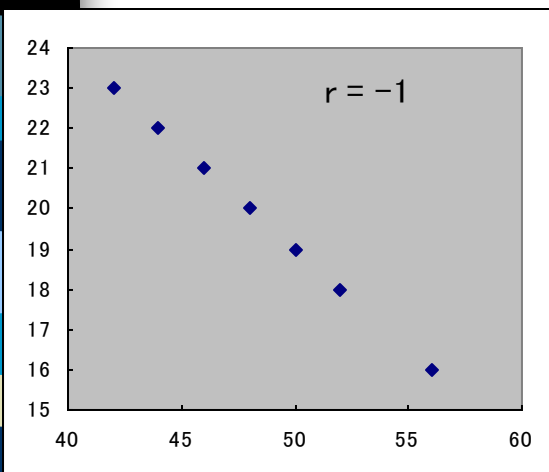
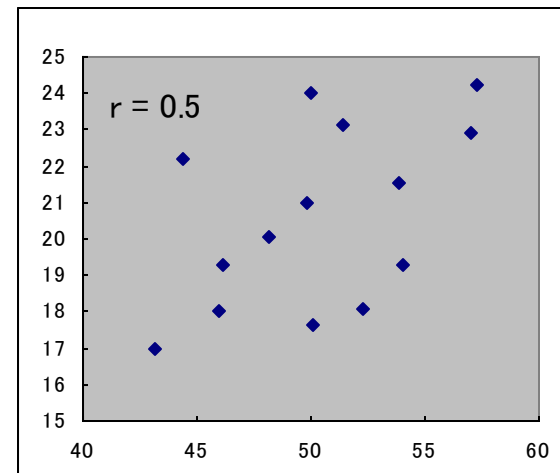
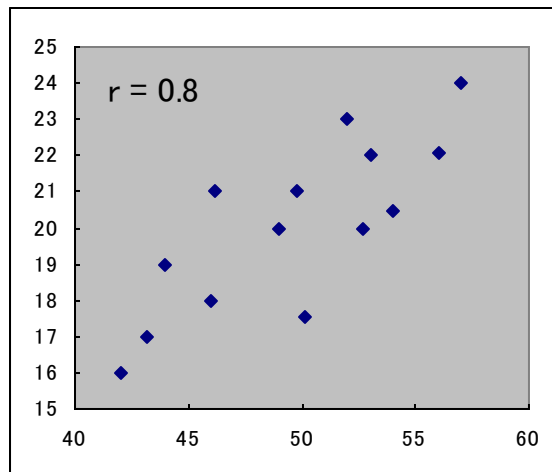
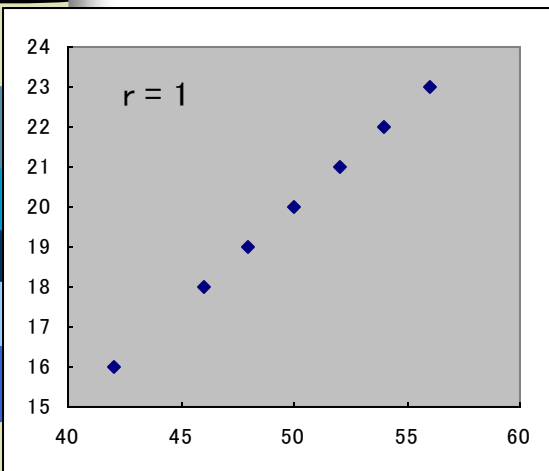
相関の強さ：絶対値  $|r|$  または  $r^2$  で評価。1に近いほど相関が強い

### ■ 相関係数

$$r = \frac{S(xy)}{\sqrt{S(xx)}\sqrt{S(yy)}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

$S(xy)$  : 偏差積和  $S(xx), S(yy)$  : 偏差平方和

# ■ 相関係数と散布図



## ■ 相関係数の算出

EXCEL

## ■ 相関係数の見方

$0.8 \leq |r|$   $\Rightarrow$  強い相関あり

$0.6 \leq |r| < 0.8$   $\Rightarrow$  相関あり

$0.4 \leq |r| < 0.6$   $\Rightarrow$  弱い相関あり

$|r| < 0.4$   $\Rightarrow$  ほとんど相関なし

これは目安、厳密には仮説検定を

## ■ 相関係数を算出する関数      CORREL

CORREL(xのデータ範囲, yのデータ範囲)



# 重要 相関係数の別解

$$r = \frac{S(xy)}{\sqrt{S(xx)}\sqrt{S(yy)}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2} \sqrt{\sum_{i=1}^n (y - \bar{y})^2}}$$

$$= \frac{\text{偏差積和}}{\sqrt{\text{偏差平方和}_x} \sqrt{\text{偏差平方和}_y}}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \bar{y})^2}} = \frac{\text{偏差積和の平均}}{\text{標準偏差}_x \times \text{標準偏差}_y}$$

$S(xy)$ : 偏差積和  $S(xx), S(yy)$ : 偏差平方和



## § 2 母相関係数に関する検定と推定

- 1 母相関係数の検定
- 2 母相関係数の推定

# 1 母相関係数の検定

【例題5-4】バドミントンのラケットの硬度(y)と重量(x)の相関係数

## ■ 考え方と適用手法

母相関係数 $\rho$ : 母集団の相関係数

相関係数  $r$  : 標本相関係数

## 無相関の検定

$r$ が0となることは、まずない。

しかし、母相関係数は0かも知れない。

⇒母相関係数 $\rho$ が0と考えられるかどうか

## ■ 無相関の検定手順

(手順1) 仮説の設定

$H_0: \rho = 0$  (母相関係数は0である)

$H_1: \rho \neq 0$  (母相関係数は0ではない)

(手順2) 有意水準 $\alpha$ の設定

有意水準 $\alpha = 0.05$

(手順3) 検定統計量  $t$  の計算 
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

(手順4)  $p$ 値の算出

$p$ 値は $t$ 分布において、 $|t|$ 以上の値が発生する確率である。

(手順5) 判定

$p$ 値  $\leq$  有意水準 $\alpha \Rightarrow$  帰無仮説 $H_0$ を棄却する

$p$ 値  $>$  有意水準 $\alpha \Rightarrow$  帰無仮説 $H_0$ を棄却しない



- 検定統計量t値と確率p値の計算 EXCEL

- 結果の見方

p値=0.0085 < 有意水準 $\alpha$ =0.05

⇒ 仮説 $H_0$ は棄却される。

⇒ 硬度と重量には相関関係が存在する

## 2 母相関係数の推定

【例題6-5】例題6-4のデータを使って、相関係数の推定

### ■ 考え方と適用手法

$\rho$ を区間推定するには、Z変換をする。

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

このZは正規分布に近似的に従う。

$$z \sim N(\mu_\rho, \sigma_\rho^2) = N \left( \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right), \frac{1}{n-3} \right)$$

## ■ 母相関係数の区間推定

(手順1) 相関係数  $r$  を算出する

(手順2)  $z$  値の算出

$$Z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

(手順3)  $z$  についての信頼区間を求める。

信頼上限

信頼下限

$$z_U = z + \frac{u(\alpha)}{\sqrt{n-3}} \quad z_L = z - \frac{u(\alpha)}{\sqrt{n-3}}$$

(手順4)  $z_U, z_L$  を  $r$  に戻して、 $\rho$  についての信頼区間は？

信頼上限

信頼下限

$$\rho_U = \frac{e^{2z_U} - 1}{e^{2z_U} + 1} \quad \rho_L = \frac{e^{2z_L} - 1}{e^{2z_L} + 1}$$

母相関係数  $\rho$  の  $(1-\alpha) \times 100\%$  の信頼区間は

$$\rho_L \leq \rho \leq \rho_U$$



■  $z \Leftrightarrow r$

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

$$2z = \ln \left( \frac{1+r}{1-r} \right)$$

$$\frac{1+r}{1-r} = e^{2z} \quad (\because y = \ln x \Leftrightarrow x = e^y)$$

$$1+r = e^{2z} - re^{2z}$$

$$(e^{2z} + 1)r = e^{2z} - 1$$

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$



### 3 相関係数のジャックナイフ推定

【例題6-6】10組のデータで1組のデータを除いた相関係数

#### ■ 考え方と適用手法 ジャックナイフ法

データを1つずつ除いた相関係数を求める。

相関係数の安定性 ⇒ はずれ値の検出

#### ■ 相関係数の算出 EXCEL

#### ■ 結果の見方

No.5のデータを除いた相関係数が大きく異なる

⇒ NO.5のデータがはずれ値

【例題6-7】例題6-6のデータで、相関係数のジャックナイフ推定値

## ■ 考え方と適用手法

データを1つずつ除いて相関係数を求め、その変化を見る。  
この相関係数  $\Rightarrow$  ジャックナイフ推定値を求める。

## ■ ジャックナイフ推定値

n個のデータ  $\Rightarrow$  母数の推定量  $g_n$

i番目のデータを除いた  $\Rightarrow$  推定量  $g_{n-1(i)}$

次のような擬似値 $\theta_i$ を定義する

$$\theta_i = ng_n - (n-1)g_{n-1(i)}$$

この擬似値 $\theta_i$ の平均値  $\Rightarrow$  ジャックナイフ推定値

ジャックナイフ推定値のねらい: 標準誤差を擬似値で把握



- ジャックナイフ推定値の算出

EXCEL

- 結果の見方

ジャックナイフ推定値 0.7342

通常のコ相関係数 0.7699

- ブートストラップ推定量

# 相関係数の応用

## POSデータの分析

- 下記のPOSデータから、いかなる販売戦略が考えられますか。

	カフェオレ 280G	十六茶34 0G缶	十六茶500 MLペット	〇〇烏龍 茶	XXチーズ	〇〇ぶた まん	◇◇カル ビ焼肉
顧客1	1	1	0	0	2	0	0
顧客2	3	0	1	0	2	0	1
顧客3	1	4	0	0	1	2	1
顧客4	0	1	0	1	0	1	1
顧客5	1	0	0	2	1	0	1
顧客6	0	1	0	0	1	1	0
顧客7	1	1	0	1	1	0	1

出所：上田データマイニング塾

[http://www.datamining.jp/jirei/j001\\_1.htm](http://www.datamining.jp/jirei/j001_1.htm)