計測工学　第６回

前回：　　・母平均の区間推定（続き）
　　　　　・相関と回帰

今回：　　・相関係数
　　　　　・回帰分析

6.1 相関係数

・例：30 人の学生における高校時の評価平均xと大学 1 年時の評価平均yの相関を考える.

表 6.1　x-y 一覧表

x	y	x	y	x	y
3.0	2.4	2.9	1.9	3.1	2.8
2.4	2.6	2.7	2.2	3.3	3.2
3.7	3.0	3.7	3.1	2.7	1.8
3.6	3.9	2.7	2.6	3.5	2.7
3.8	3.6	3.3	2.8	2.9	2.1
2.9	3.0	2.8	2.7	2.7	1.7
3.5	3.1	3.1	2.4	2.9	1.7
3.0	2.8	2.8	3.0	3.2	2.3
2.3	2.2	3.0	3.3	3.4	2.6
3.0	2.9	2.2	1.8	2.5	2.7

① 散布図を作成

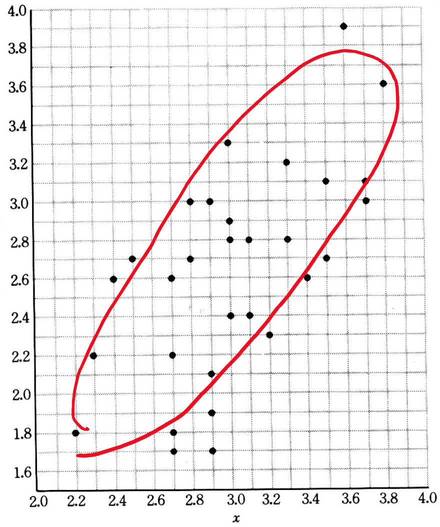

図 6.1　x-y 散布図

② 変数(x_i, y_i)に対して、

$$x' = \frac{x_i - \bar{x}}{S_x}, \quad y_i = \frac{y_i - \bar{y}}{S_y}$$

による標準化を行い、再プロットする。

ここで $S_x^2 = S_{xx} = \frac{1}{n}\sum (x_i - \bar{x})^2$: xの標本分散

$S_x = \sqrt{S_x^2}$

$S_y^2 = S_{yy} =$ 〃 $(y_i - \bar{y})^2$: yの 〃

$S_{xy} = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$: xとyの共分散

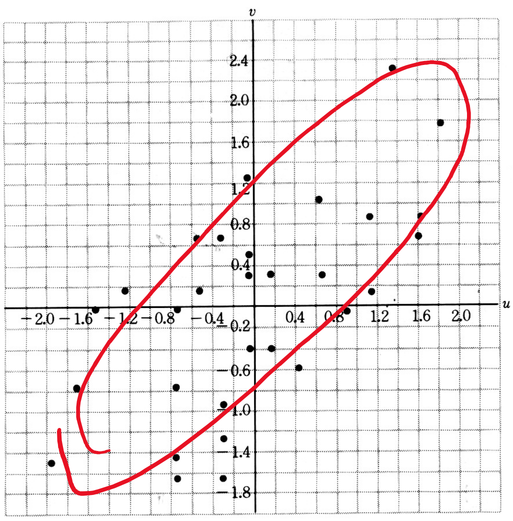

図 6.2　標準化 x-y 散布図

③ x'とy'の積和の平均 $\frac{1}{n}\sum x'y'$ を「相関係数r」と定義する。

$$r = \frac{1}{n}\sum\left\{\frac{x_i - \bar{x}}{S_x}\cdot\frac{y_i - \bar{y}}{S_y}\right\} = \frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{S_x \cdot S_y}$$

$$= \frac{S_{xy}}{\sqrt{S_x^2}\cdot\sqrt{S_y^2}}$$

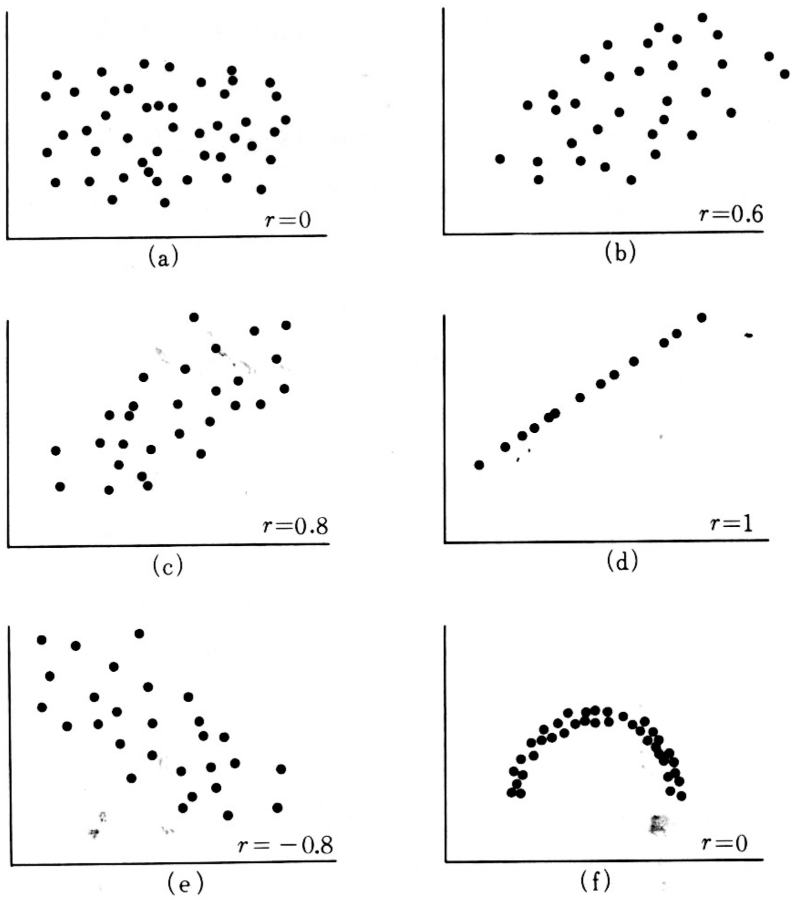

図 6.3　r とデータの傾向の関係

$r = 0$: 相関無し

$r > 0$: 正の相関 =右肩上がり

$r < 0$: 負の相関 =右肩下がり

$0 \leq |r| \leq 0.2$: ほとんど相関無し

$0.2 < |r| \leq 0.4$: 弱い相関

$0.4 < |r| \leq 0.7$: ある程度の相関

$0.7 < |r| \leq 1.0$: 強い相関

左図(f)より:
この相関係数rは
線形関係を判定
しているため.

・例題: 学生の入学試験順位 x と 1 年後の順位 y を調べ, そのうち 5 名については (x, y) →
(10, 40), (25, 35), (35, 10), (60, 50), (70, 70) であった. x と y の相関係数を求めよ.

No.	1	2	3	4	5	
x_i	10	25	35	60	70	$\bar{x}$
$x_i - \bar{x}$	−30	−15	−5	20	30	40
y_i	40	35	10	50	70	$\bar{y}$
$y_i - \bar{y}$	−1	−6	−31	9	+29	41

$$S_x = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})}$$

$S_x^2 = 490$

$$= \sqrt{\frac{1}{5}\{(-30)^2 + (-15)^2 + (-5)^2 + 20^2 + 30^2\}} = 22.13\cdots$$

$$S_y = \sqrt{\frac{1}{5}\{(-1)^2 + (-6)^2 + (-31)^2 + 9^2 + 29^2\}} = 19.59\cdots$$

$S_y^2 = 384$

$$S_{xy} = \frac{1}{5}\{(-30\times -1) + (-15\times -6) + (-5\times -31) + (20\times 9) + (30\times 29)\}$$
$$= 265$$

$$\therefore r = \frac{S_{xy}}{S_x \cdot S_y} = 0.611\cdots = \underline{0.61}$$

$$r = \frac{S_{xy}}{\sqrt{S_x^2}\cdot\sqrt{S_y^2}} = 0.6109\cdots$$
$$= 0.61$$

6.2 回帰分析

●最小 2 乗法による線形回帰分析

・回帰式として $y = ax + b$ を想定する：データから如何に a と b を求めるか？

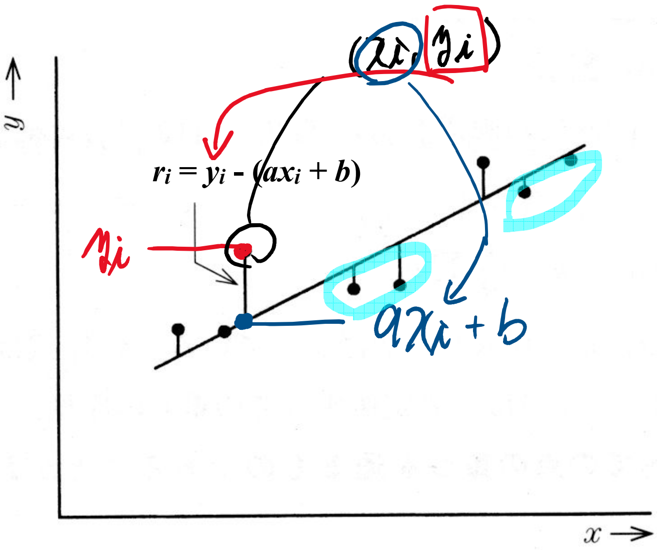

図 6.4 最小 2 乗法による線形回帰分析

・ n 個の測定値の対 (x_1, y_1), (x_2, y_2)・・・(x_n, y_n)

を対象とする

・x に誤差はなく，誤差は全て y に含まれると仮定

⇩

・ある傾き a と切片 b を持つ直線 $f(x) = ax + b$ と，

i 番目の測定値 y_i との残差 r_i

$\rightarrow$ $r_i = y_i - f(x_i) = y_i - (ax_i + b)$

⇩

・ n 個の測定値に対して，残差 r_i の 2 乗和 $\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \{y_i - (ax_i - b)\}^2$ が最小となるのが

最良の回帰直線

・問い：なぜ残差 r_i に対して 2 乗和を取るのか？ 残差に生じる符号の影響を排除するため．

・$\sum_{i=1}^{n} \{y_i - (ax_i - b)\}^2 = R$ とおいたとき・・・ x_i, y_i ：定数（測定値）

a, b ：変数（回帰式の係数として決定する未知数）

⇩

・ R が最小になるような a と b を決定する．

$$R=\sum_{i=1}^{n}(y_i-ax_i-b)^2=\sum_{i=1}^{n}\left(y_i^2+a^2x_i^2+b^2-2by_i-2ax_iy_i+2abx_i\right)$$

$$=\underbrace{\sum_{i=1}^{n}y_i^2}_{A}+a^2\underbrace{\sum_{i=1}^{n}x_i^2}_{B}+nb^2-2b\underbrace{\sum_{i=1}^{n}y_i}_{C}-2a\underbrace{\sum_{i=1}^{n}x_iy_i}_{D}+2ab\underbrace{\sum_{i=1}^{n}x_i}_{E}$$

$$=A+a^2B+nb^2-2bC-2aD+2abE$$

⇩

・$\boldsymbol{R}$ を変数 $\boldsymbol{a}$, $\boldsymbol{b}$ で偏微分して 0 と置く・・・最小となる条件！

$$\left.\begin{aligned}\frac{\partial R}{\partial a}&=2aB-2D+2bE=0\\ \frac{\partial R}{\partial b}&=2nb-2C+2aE=0\end{aligned}\right\}\quad a=\frac{nD-CE}{nB-E^2},\quad b=\frac{BC-DE}{nB-E^2}$$

$$a=\frac{n\Sigma x_iy_i-\Sigma x_i\Sigma y_i}{n\Sigma x_i^2-(\Sigma x_i)^2},\quad b=\frac{\Sigma x_i^2\Sigma y_i-\Sigma x_iy_i\Sigma x_i}{n\Sigma x_i^2-(\Sigma x_i)^2}$$

・例題：P.2 の例題［学生の入学試験順位 x と 1 年後の順位 y, $\bar{x}=40$, $\bar{y}=41$］について最小二乗法による線形回帰分析を行え.

$n=5$

$$\Sigma x_i=n\cdot\bar{x}=200,\quad \Sigma y_i=n\cdot\bar{y}=205$$

$$\Sigma x_i^2=(10^2+25^2+35^2+60^2+70^2)=10450$$

$$\Sigma x_iy_i=(10\times40+25\times35+35\times10+60\times50+70\times70)$$
$$=9525$$

$$a=\frac{5\cdot9525-200\cdot205}{5\cdot10450-(200)^2}=0.5408\cdots=\underline{0.54}$$

$$b=\frac{10450\cdot205-9525\cdot200}{5\cdot10450-(200)^2}=19.36\cdots=\underline{19}$$

$$\underline{y=0.54x+19}$$

6.3　第 6 回講義に関する意見･感想･質問のまとめ

●意見・感想

・最小二乗法についての説明がわかりやすかった，回帰分析は間違えそうなので実際はコンピュータに任せたい，最小二乗法は実験の時苦戦したが改めて学べてよかった，回帰分析の原理がわかってよかった，回帰分析の大変さがわかったので Excel に感謝する，回帰分析の計算方法を知らなかったので学びになった，計算は大変だったが仕組みを理解することができた：17←本当に Excel の偉大さを思い知ります．

・相関係数について重点的に復習する，覚えなければならないものが増えてきたので復習する，テストが近いのでしっかり勉強する，回帰分析の流れを復習する：10

・電卓の機能を使って楽に計算できることが実感できた，今回の小テストで初めて関数電卓の統計を使ったが煩雑な計算がすぐにできた，関数電卓の機能を確認する，電卓を使いこなせるようになってきた，相関係数の例題でデータ入力を間違えてしまった，たくさんのデータを入力するため間違ってしまう：10←どんなものでも，少しでも使いこなせればだいぶ楽になることは多々あります．

・小テストで有効数字をミスってしまった，小テスト上手くできた気がする．小テストが少しわかった，式を覚えていたので上手く行った，計算の手順をしっかり理解すれば難しくはない，u と u^2 を混同してしまった，復習が足りておらず上手くいかなかった：8←今回の小テストは平均 7.6 点，満点 28 名でした．変わらず良い結果だと思います．

・式が複雑化して頭で理解が難しかった，後半は式が複雑で混乱した，最小二乗法の例題で計算を間違えた，回帰分析の式の記号が多くて頭がゴチャゴチャした，最近少し複雑で計算が難しい，かなり複雑な指揮を計算したので疲れた，そこまで難しい内容ではないはずなのに漢字や式の羅列で惑わされた：7

・高校生の頃を思い出した，高校の時相関係数の計算をやった，ほとんど知ってる内容だった：3

・明日の深夜の日本代表の試合が待ち遠しい，明日はブラジル戦なので早く寝て備えたい：2←本当に悩ましいですね･･･明日は 1 限あるし･･･

・標本数が 5 程度であれば相関係数を求めるのは難しくなさそう←でも入力ミスが怖いですよね．

・相関係数について理解が深まった：2

・わからないところはミニッツペーパーの質問への回答として説明してくれるので質問しやすい←そういってもらえると回答してる甲斐があります．

・授業早く終わって嬉しい

●質問

・(授業とは離れるが)就活した時何を重視したのか，なぜ大学の先生になったのか？←自動車メーカーに行きたかったんですがバブル崩壊で全然無理そうだったので代わりにドクターに行った，って感じです(^^;

・応用数理 E を参考にしろとのことだが教員同士で科目内容を共有しているのか？←してません．

・最小二乗法で求める回帰式の切片 b の有効数字の取り扱いは？←この問題の場合，与えられた値も測定値ではないし，回帰式の予測の対象が「順位」なので，最小桁が 1 の位になるように合わせました．

・1 ページ目の y の標準化は yi ではなくて y'では？←ご指摘ありがとうございます，下記間違いです．修正しました．

・中間テストの内容や対策として何をすればいいのか？←まずは授業中の例題や小テストの問題を完全に理解し解けるようにしておくことですね．

・標本分散なのに 1/(n-1)ではないのか？←正直私も不思議に思っていたのですが，質問をもらったので調べた結果，式としては標本分散ー標本共分散のセットか不偏分散ー不偏共分散のセットになり，1/n もしくは 1/(n-1)の係数は分母と分子で相殺されるから，どちらでも良いとのことでした．

6.4 第 5 回小テスト解答

Q.1　正規分布する母集団（σ^2 未知）から以下の標本を取出したとき，母平均 μ の信頼区間を 68.3%の信頼係数で求めよ．計算過程も明示すること．　　　標本：12.9, 13.2, 13.1, 12.8

A.1　標本平均 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = 13.0，標本数 $n = 4$

不偏分散 $u^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.0333\ldots$

t 分布表より　$t_{n-1}(Z\%) = 3.182$

$t_{n-1}(Z\%)\sqrt{\frac{u^2}{n}}$ = 0.290... = 0.3

∴ $12.7 \leqq \mu \leqq 13.3$

自由度 $(N-1)$	標本数 (N)	$t_{N-1}(z\ \%)$，$z(\%)$：信頼度 68.3%	90%	95%	99%
		P（有意水準），両側検定 31.7%	10%	5%	1%
1	2	1.837	6.314	12.706	63.657
2	3	1.321	2.920	4.303	9.925
3	4	1.197	2.353	3.182	5.841
4	5	1.141	2.132	2.776	4.604
5	6	1.110	2.015	2.571	4.032
6	7	1.090	1.943	2.447	3.707
7	8	1.077	1.895	2.365	3.500
8	9	1.066	1.860	2.306	3.355
9	10	1.059	1.833	2.262	3.250