データ解析試験問題
(持ち込み不可, 時間 10:25-11:50)

作成: 安藤 和敏 (静岡大学工学部)

2003 年 2 月 12 日

試験についての注意

- A〜D までの全ての問題を解きなさい.
- 電卓と筆記用具以外のものは机の下にしまいなさい.
- 数値の計算結果は, (小数点以下第 3 位を四捨五入して) 第 2 位までを記すこと.
- 証明問題に対する答案はなるべく詳しく書いてあるものを評価する.
- 11 時 25 分以降, 退室を許可する. 他の人の迷惑にならないように退室すること.
- 問題用紙は持ち帰ってよい.
- 解答は, Web ページで本日中に公開する.

A (計 37 点)

(i) (各 4 点)

変数 x に関する n 個のデータ $x_1, x_2, \cdots, x_n$ が与えられているときに, これらのデータの平均 $\bar{x}$ と分散 s_{xx} は,

$$\bar{x} = \boxed{(1) \qquad\qquad\qquad\qquad},$$

$$s_{xx} = \boxed{(2) \qquad\qquad\qquad\qquad}$$

である. さらに, 変数 y に関する n 個のデータ $y_1, y_2, \cdots, y_n$ が与えられているときに, x と y の共分散 s_{xy} は

$$s_{xy} = \boxed{(3) \qquad\qquad\qquad\qquad}$$

である.

表 1: 変数 x, y のデータ

i	1	2	3	4	5	6	7	8	9	10
x_i	53	56	47	15	53	55	76	12	39	41
y_i	71	75	63	22	71	74	101	18	53	56

(ii) (各 3 点)

x_i, y_i $(i = 1, \ldots, n)$ が以下の表 1 のように与えられているとする．問題 A(i) の答え (1)〜(3) に基づいて, $\bar{x}$, $\bar{y}$, s_{xx}, s_{yy}, s_{xy} を計算すると

$$\begin{aligned} \bar{x} &= \boxed{(4)\qquad\qquad\qquad} \\ \bar{y} &= 60.4 \\ s_{xx} &= \boxed{(5)\qquad\qquad\qquad} \\ s_{yy} &= 562.44 \\ s_{xy} &= \boxed{(6)\qquad\qquad\qquad} \end{aligned}$$

となる．

(iii) ((8) と (9) は 5 点．その他は 3 点)

表 1 の x と y の間に

$$y = ax + b$$

という線型の関係があると仮定して, 傾き a と b を求めることを考えよう．誤差項 e_i を加えた

$$y_i = ax_i + b + e_i \quad (i = 1, \cdots, n)$$

というモデルを $\boxed{(7)\qquad\qquad\qquad}$ と呼ぶ．誤差の 2 乗和を

$$F(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

とおく．$F(a, b)$ を最小にする a と b を, それぞれ, $\hat{a}, \hat{b}$ とすると, $(\hat{a}, \hat{b})$ は連立方程式

$$\begin{aligned} \frac{\partial F(a, b)}{\partial a} &= 0, \\ \frac{\partial F(a, b)}{\partial b} &= 0 \end{aligned}$$

の解である．$F(a, b)$ を a と b で, それぞれ, 偏微分すると,

$$\begin{aligned} \frac{\partial F(a, b)}{\partial a} &= \boxed{(8)\qquad\qquad\qquad}, \\ \frac{\partial F(a, b)}{\partial b} &= -2 \sum_{i=1}^{n} \{y_i - (b + ax_i)\} \end{aligned}$$

であるから, $\hat{a}$ と $\hat{b}$ を $\bar{x}, \bar{y}, s_{xx}, s_{yy}, s_{xy}$ を用いて表すと,

$$\hat{a} = \frac{s_{xy}}{s_{xx}}, \quad \hat{b} = \boxed{(9) \qquad\qquad\qquad\qquad}$$

となる. 問題 A(ii) で求めた数値から, $\hat{a}$ と $\hat{b}$ を計算すると,

$$\hat{a} = 1.2948928 \ , \quad \hat{b} = \boxed{(10) \qquad\qquad\qquad\qquad}$$

を得る.

B　(計 22 点)

問題 A で考えたモデル

$$y_i = ax_i + b + e_i \quad (i = 1, \cdots, n)$$

において, e_i $(i = 1, \cdots, n)$ が互いに独立に, 同一の正規分布 $N(0, \sigma^2)$ に従う確率変数であると仮定する. ここで, 分散 σ^2 は未知のパラメータである.

このとき, 各 $i = 1, \cdots, n$ について, y_i は期待値が $ax_i + b$ で分散が σ^2 のの正規分布に従う. さらに, $i \neq j$ に対して, y_i と y_j は独立である. a の推定量 $\hat{a}$ は,

$$\hat{a} = \frac{s_{xy}}{s_{xx}} = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{ns_{xx}} \right) y_i$$

と表される. $\hat{a}$ は, 正規分布に従う確率変数 y_i $(i = 1, \cdots, n)$ の線型結合で表されているので, 正規分布に従う.

(i)　(10 点)

$\mathrm{E}(\hat{a}) = a$ を証明しなさい.

(ii)　(2 点)

$\hat{a}$ のように, その期待値が真の値と一致するような推定量は何と呼ばれるか?

(iii)　(10 点)

$\mathrm{V}(\hat{a}) = \frac{\sigma^2}{ns_{xx}}$ を証明しなさい.

C　(問題 B の続き) (計 20 点)

問題 B の (1) と (3) より, $\hat{a}$ は正規分布 $N\left(a, \frac{\sigma^2}{ns_{xx}}\right)$ に従う確率変数であるから, 標準化した

$$u = \frac{\hat{a} - a}{\sqrt{\frac{\sigma^2}{ns_{xx}}}}$$

は, 標準正規分布 $N(0,1)$ に従う. さらに, 上式の σ^2 を

$$V_e = \frac{F(\hat{a},\hat{b})}{n-2}$$

で置き換えた,

$$t = \frac{\hat{a}-a}{\sqrt{\frac{V_e}{ns_{xx}}}}$$

は, 自由度 $n-2$ の t 分布に従うことが知られている.

(i)　(8 点)

以下の等式を証明せよ.

$$F(\hat{a},\hat{b}) = n\left(s_{yy} - \frac{s_{xy}^2}{s_{xx}}\right).$$

(ii)　(4 点)

表 1 のデータに対する V_e の値を計算せよ.

(iii)　(8 点)

$t_{.05}(n-2)$ を t 分布の 5%限界値とするとき, a の 95%信頼区間を式で表せ.

D　(21 点)

図 1 のようなデータが与えられたときに, 線型重回帰モデル

	A	B	C	D	E
1		y	x1	x2	x3
2	1	4	13	24	63
3	2	35	16	34	54
4	3	71	27	40	51
5	4	88	25	58	57
6	5	56	23	26	36
7	6	46	25	32	60
8	7	32	16	20	36
9	8	60	22	32	39
10	9	59	19	44	51
11	10	30	11	52	75

図 1: 入力データ

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} + e_i \quad (i = 1, \cdots, 10).$$

を立てて分析を行うことを考えよう．ここで, e_i $(i = 1, \cdots, 10)$ は互いに独立に同一の分布 $N(0, \sigma^2)$ に従う確率変数である.

このデータに対して, エクセルの分析ツールで回帰分析を実行すると, 図 2 のような出力が得られた.

(i)　(2 点)

回帰係数 $\hat{a}_0$ の数値はいくつか?

(ii)　(2 点)

a_0 の 95%信頼区間を数値で表せ.

(iii)　(5 点)

$V_e = \frac{F(\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3)}{10-3-1}$ の数値いくつか?

(iv)　(6 点)

V_e と誤差の分散 σ^2 は, どのような関係にあるかを述べよ.

(v)　(6 点)

仮説 $a_0 = 3.0$ を検定しなさい.

概要

回帰統計	
重相関 R	0.9999469088
重決定 R 2	0.9998938203
補正 R 2	0.9998407305
標準誤差	0.3017990782
観測数	10

分散分析表

	自由度	変動	分散	測された分散比	有意 F
回帰	3	5146.353504	1715.45117	18833.998959	2.618513e-12
残差	6	0.546496102	0.09108268		
合計	9	5146.9			

	係数	標準誤差	t	P-CM	下限 95%	上限 95%
切片	1.9486383196	0.736841763	2.64458181	0.0383050273	0.1456501592	3.75162648
X CM 1	2.2141658684	0.022246454	99.5289339	6.932939e-11	2.1597307162	2.2686010207
X CM 2	1.500413809	0.011046947	135.821582	1.074287e-11	1.4733828839	1.5274447341
X CM 3	-0.992005475	0.011573656	-85.712371	1.698682e-10	-1.020325211	-0.963685739

図 2: 回帰分析の結果

解答用紙

学籍番号: ______________________

氏名: ______________________

A

(1) ______________________ (2) ______________________

(3) ______________________ (4) ______________________

(5) ______________________ (6) ______________________

(7) ______________________ (8) ______________________

(9) ______________________ (10) ______________________

B

(i)

(ii)

(iii)

C

(i)

(ii) ______________________________

(iii) ____

D

(i) ______________________________

(ii) ______________________________

(iii) ______________________________

(iv) ____

(v) ____