データ解析 2004 年度中間試験問題

静岡大学工学部システム工学科

安藤 和敏

2004 年 12 月 15 日

注意事項

- A4 片面に手書きのカンニングペーパー, 及び, (関数) 電卓のみ持ち込み可.
- 印刷あるいはコピーしたカンニングペーパーやノートパソコンの持ち込みは不可.
- 携帯電話の電卓機能の使用ももちろん不可.
- 試験の時間は 10:20-11:40 である.
- 問題用紙は持ち帰ってよい.

問題 1．(配点 18)

x_i $(i = 1, \cdots, n)$ が正規分布 $N(\mu, \sigma^2)$ にしたがう確率変数であるとする．ここで, μ と σ^2 は未知のパラメータである．さらに, $i \neq j$ に対して x_i と x_j は独立であるとする．

以下の事実を用いて, μ の信頼区間を求めてみよう．

(i) 正規分布にしたがう確率変数の線形結合は正規分布にしたがう確率変数である．

(ii) 正規分布にしたがう確率変数を標準化すると標準正規分布にしたがう確率変数となる．

(iii) 任意の n 個の確率変数 y_i $(i = 1, \cdots, n)$ と定数 a_i $(i = 1, \cdots, n)$ に対して $E(\sum_{i=1}^n a_i y_i) = \sum_{i=1}^n a_i E(y_i)$.

(iv) 任意の n 個の独立な確率変数 y_i $(i = 1, \cdots, n)$ と定数 a_i $(i = 1, \cdots, n)$ に対して $V(\sum_{i=1}^n a_i y_i) = \sum_{i=1}^n a_i^2 V(y_i)$.

設問 (1) 以下の文章中の空欄 $\boxed{\text{ア}}$, $\boxed{\text{イ}}$, $\boxed{\text{エ}}$, $\boxed{\text{カ}}$ の中にはもっとも良くあてはまる命題を (i)〜(iv) から選び, これら以外の空欄には最も適切な数式または数値を記入せよ．

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

は, $\boxed{\text{ア}}$ より, 正規分布にしたがう確率変数になる．$\bar{x}$ の期待値は, $\boxed{\text{イ}}$ によって $\boxed{\text{ウ}}$ であり, $\bar{x}$ の分散は, $\boxed{\text{エ}}$ によって $\boxed{\text{オ}}$ であるから, $\bar{x} \sim N(\boxed{\text{ウ}}, \boxed{\text{オ}})$ である．

$\boxed{\text{カ}}$ によって, $\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ を得る．ここで, $\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$ において, σ^2 のかわりに V_x を代入した $\frac{\bar{x}-\mu}{\sqrt{V_x/n}}$ は, 自由度 $\boxed{\text{キ}}$ の t-分布にしたがう確率変数となる．

したがって, $t(\phi, 0.05)$ を自由度 ϕ の t 分布の両側 5%点とすると,

$$-t(\boxed{\text{キ}}, 0.05) \leq \frac{\bar{x} - \mu}{\sqrt{V_x/n}} \leq t(\boxed{\text{キ}}, 0.05)$$

となる確率は, $\boxed{\text{ク}}$ である．

上式を書き直せば, μ の信頼率 $\boxed{\text{ク}}$ の信頼区間は, $\boxed{\text{ケ}}$ であることがわかる．

問題 2. (配点 20)

表 1 に示したデータは, 正規分布 $N(\mu, \sigma^2)$ にしたがう母集団から採取された $n = 10$ 個のデータである. ここで, μ と σ^2 は未知のパラメータである.

表 1:

No	1	2	3	4	5	6	7	8	9	10
x	48	37	40	39	48	44	46	51	50	40

設問 (1) 帰無仮説 $H_0 : \mu = 40$, 対立仮説 $H_1 : \mu \neq 40$ として, 有意水準 5%で検定を行なえ.

設問 (2) μ の 95%信頼区間を (数値で) 求めよ.

問題 3. (配点 32)

表 3 に示すような 2 変数 x, y についてのデータが得られているとする.

表 2:

No	x	y
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
i	x_i	y_i
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

設問 (1) 以下の文章中の空欄 $\boxed{\text{ア}}$ ～ $\boxed{\text{エ}}$ の中に最も適切な数式または語句を記入せよ.

最小 2 乗法は, データ (x_i, y_i) に最も良くあてはまる直線を求める方法の一つである. ここで, 直線 $y = \hat{\beta}_0 + \hat{\beta}_1 x$ のあてはまりの度合を測る尺度として, 残差平方和

$$S_e = \boxed{\text{ア}}$$

が用いられる. この尺度に基づけば, データ (x_i, y_i) に最も良くあてはまる直線を求めるためには, S_e を最小にする $\hat{\beta}_0$ と $\hat{\beta}_1$ を求めればよい. すなわち, 次の連立方程式の解を求めればよい.

$$\begin{aligned}\frac{\partial S_e}{\partial \hat{\beta}_0} &= -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0,\\ \frac{\partial S_e}{\partial \hat{\beta}_1} &= -2\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.\end{aligned}$$

$\hat{\beta}_0$,$\hat{\beta}_1$ について整理すると,

$$\begin{aligned}\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i,\\ \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i\end{aligned}$$

となる. この連立方程式は, $\boxed{\text{イ}}$ と呼ばれる. $\boxed{\text{イ}}$ を解いて,

$$\begin{aligned}\hat{\beta}_0 &= \boxed{\text{ウ}},\\ \hat{\beta}_1 &= \boxed{\text{エ}}\end{aligned}$$

を得る.

設問 (2) 設問 (1) の文章中に現れた式を用いて, $\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y}\} = 0$ を証明せよ.

設問 (3) $\sum_{i=1}^{n} e_i\{(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y}\} = 0$ を用いて, $S_{yy} = S_e + \sum_{i=1}^{n}\{(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{y}\}^2$ を証明せよ.

問題 4. (配点 30)

表 3 に示すような 2 変数 x, y についての $n = 10$ 個のデータが得られているとする.

表 3:

No	x	y
1	69	-147
2	43	-76
3	32	-53
4	40	-72
5	68	-141
6	43	-79
7	38	-58
8	49	-95
9	34	-53
10	62	-124

以下の設問に答えよ.

設問 (1) 回帰直線 $y = \hat{\beta}_0 + \hat{\beta}_1 x$ の y 切片 $\hat{\beta}_0$ 及び傾き $\hat{\beta}_1$ を数値で求めよ.

設問 (2) 寄与率 R^2, 及び, 自由度調整済み寄与率 R^{*2} を数値で与えよ. また, それらの値から, なにが言えるか?

学籍番号		氏名	

問題 1(1) の解答欄

ア. ______________________ イ. ______________________

ウ. ______________________ エ. ______________________

オ. ______________________ カ. ______________________

キ. ______________________ ク. ______________________

ケ. ____

問題 2(1) の解答欄

問題 2(2) の解答欄

学籍番号		氏名	

問題 3(1) の解答欄

ア. ____________________

イ. ____________________

ウ. ____________________

エ. ____________________

問題 3(2) の解答欄

問題 3(3) の解答欄

学籍番号		氏名	

問題 4(1) の解答欄

問題 4(2) の解答欄