統計検定
Japan Statistical Society Certificate

1級
統計応用

2024 年 11 月 17 日

【注意事項】

1 **試験開始の合図があるまで，この問題冊子の中を見てはいけません。**

2 この問題冊子は，60 ページあります。**3〜13 ページ**は**人文科学**，**15〜27 ページ**は**社会科学**，**29〜39 ページ**は**理工学**，**41〜51 ページ**は**医薬生物学**の問題です。

3 選択した分野の問題 5 問から 3 問を選択して，解答しなさい。

4 試験時間は 90 分です。

5 試験中に問題冊子の印刷不鮮明，ページの落丁・乱丁および解答冊子の汚れ等に気付いた場合は，手を挙げて監督者に知らせなさい。

6 **解答冊子・マークシートの A 面には次の項目があるので，それぞれの指示に従い記入あるいは確認しなさい。**項目の内容に誤りがある場合は，手を挙げて監督者に知らせなさい。

① **マークシート**の**氏名**
氏名を**記入**しなさい。

② **マークシート**の**検定種別**と**受験番号**
受験する検定種別と**受験番号**を確認しなさい。

③ **マークシート**の **Web 合格発表**
Web 合格発表について，希望の有無を**マーク**しなさい。

④ **解答冊子表紙**の左上に**受験番号**を**記入**し，マーク欄の該当する**数字**を**塗りつぶしなさい**。

⑤ **問**ごと（**問 1**，**問 2**，...）に解答のページを改めなさい。
なお，**解答する問の順番**（**問 3**，**問 1**，**問 5** など）は**問いません**。

⑥ 各ページの先頭に受験番号と問題番号を書きなさい。（この冊子裏面の記入例参照）
解答冊子には，最終的な解答だけでなくそれに至る過程も記しなさい。最終的な解答が正しくないときにも点が与えられることがあります。

⑦ 解答冊子の表紙の選択分野欄の選択した分野と，問題番号欄の選択した問題番号をそれぞれ ◯ で囲みなさい（得点欄には何も書かないこと）。（この冊子裏面の記入例参照）

7 53 ページ以降に付表を掲載しています。必要に応じて利用しなさい。

8 問題冊子の余白等は適宜利用してよいが，どのページも切り離してはいけません。

9 試験終了後，問題冊子は持ち帰りなさい。

（冊子裏面につづく）

人文科学

人文科学

問 1　A 高校で行われた 5 科目（国語，英語，社会，数学，理科）の試験得点間の関係を主成分分析によって分析することとした。各科目の得点を表す変数を上述の科目順に x_1, x_2, x_3, x_4, x_5 として，以下の各問に答えよ。

なお，得点の標準化とは，得点 x の平均を $\bar{x}$，標準偏差を s としたとき，$y = (x - \bar{x})/s$ と変数変換して平均を 0，標準偏差を 1 にすることをいい，そのときの y を標準化得点と呼ぶ。

〔1〕　表 1 は，試験得点の科目別の平均と標準偏差および科目間の相関係数である。

表 1：試験の科目別得点の概要

科目	国語 (x_1)	英語 (x_2)	社会 (x_3)	数学 (x_4)	理科 (x_5)
平均	66.0	64.5	66.5	67.5	69.0
標準偏差	19.7	19.2	14.8	19.1	16.4

相関係数	国語 (x_1)	英語 (x_2)	社会 (x_3)	数学 (x_4)	理科 (x_5)
国語 (x_1)	1.00	0.95	0.89	0.32	0.35
英語 (x_2)	0.95	1.00	0.94	0.41	0.47
社会 (x_3)	0.89	0.94	1.00	0.64	0.66
数学 (x_4)	0.32	0.41	0.64	1.00	0.98
理科 (x_5)	0.35	0.47	0.66	0.98	1.00

〔1-1〕　国語 (x_1) と数学 (x_4) の得点で作る 2×2 の相関行列 R_2 の固有値 λ_1,λ_2 $(\lambda_1 > \lambda_2)$ と，それらに対応する長さ 1 の固有ベクトル $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ を，表 1 の数値を用いて具体的に求めよ。

〔1-2〕　国語と数学の得点を標準化し，それぞれ u_1, u_4 とする。これらを用いて主成分分析を行い，第 1 主成分を $z_1 = a_1u_1 + a_2u_4$，第 2 主成分を $z_2 = b_1u_1 + b_2u_4$ と置いた。上問〔1-1〕で求めた値から，第 1 主成分および第 2 主成分の各寄与率と，第 1 主成分 z_1 と国語の標準化得点 u_1 の相関係数である主成分負荷量 $\rho_{1,1}$ を求めよ。

〔2〕　5科目の得点全体を標準化したうえで主成分分析を行った。表2は，第1主成分から第3主成分までの固有値と固有ベクトルである。

表2：固有値と固有ベクトル（第3主成分まで）

固有値	第1主成分	第2主成分	第3主成分
	3.67	1.24	0.05

固有値ベクトル	第1主成分	第2主成分	第3主成分
国語	−0.44	−0.46	−0.69
英語	−0.47	−0.37	0.53
社会	−0.51	−0.15	0.21
数学	−0.40	0.58	−0.38
理科	−0.41	0.54	0.24

〔2-1〕　表2の固有値から，第1主成分の寄与率，および第2主成分までの累積寄与率を求めよ。また，固有ベクトルの値から，第1主成分と第2主成分を解釈せよ。

〔2-2〕　5科目の試験得点は5変量正規分布に従うとし，標本統計量を対応する母集団パラメータと見なして，第1主成分得点の上側20%に相当する生徒の第1主成分得点を求めよ。また，標準化得点を合計した合計得点の分散が18.24であるとき，第1主成分得点と合計得点との間の相関係数を求めよ。

〔3〕　図1は，10名の生徒に対する第1主成分得点（横軸）と第2主成分得点（縦軸）をプロットしたものである。標準化得点について，国語が1.22，英語が1.33，社会が1.58，数学が1.44，理科が1.28である生徒がいる。この生徒は図1の何番の位置にあるかを示し，その理由を述べよ。

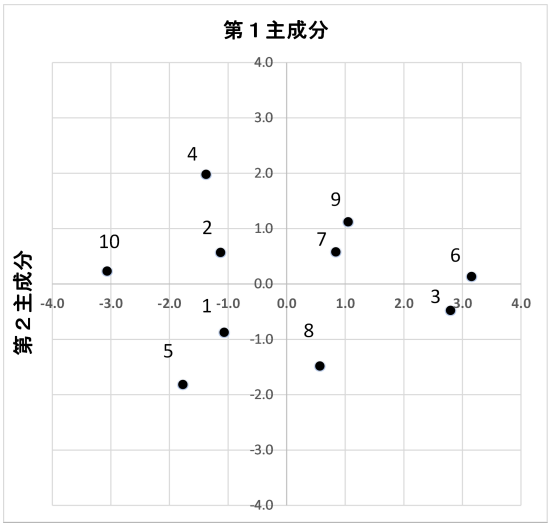


図1：第1主成分と第2主成分

問 2　ある大学の入学試験では，午前に第 1 科目を受験し，午後に第 2 科目を受験する。午前の試験の点数 X と，午後の試験の点数 Y に対し，合計点 $X+Y$ が 120 点以上を合格，というのが現在の合格基準である。この基準について教員Aと教員Bが議論をしている。次の文中の【ア】〜【ケ】の数値を求め，解答用紙に記せ。

なお，標準正規分布 $N(0,1)$ に従う確率変数を Z とするとき，$E[Z \mid Z \geq 0] = \sqrt{2/\pi}$ であることを用いてもよい。

A： 最近の試験の状況から合格率の予測結果を報告してくれない？

B： 2 つの試験の点数 $\begin{pmatrix} X \\ Y \end{pmatrix}$ は 2 変量正規分布 $N\left(\begin{pmatrix} 50 \\ 50 \end{pmatrix}, \begin{pmatrix} 10^2 & 60 \\ 60 & 10^2 \end{pmatrix}\right)$ に従っていると考えてよいです。このとき，合計点の期待値 $E[X+Y]$ が【ア】で，分散 $V[X+Y]$ が【イ】ですね。120 点以上が合格なので，合格率は【ウ】と計算できます。

A： ちょっと少ないな。合格率を 20%くらいにしたいのだけどどうだろうか？

B： 合格率を 20%にするには合格最低点を【エ】にすることになりますよ。

A： それほど大きな差はないから次回の会議で提案してみよう。

B： ここ 2 年，受験者が急増しているので，午前の試験の点数の 50 点未満を足切りして，午後の試験は午前の試験の点数上位の約半数だけを採点するのはどうですか？

A： ちょっと議論を要する考えだと思うけど... ところで，午前の試験が 50 点の人は，現在どのくらい合格していると考えていいのかな？

B： 午前の試験が 50 点の人の条件付き期待値 $E[Y \mid X = 50]$ と条件付き分散 $V[Y \mid X = 50]$ はそれぞれ【オ】と【カ】なので，$P(X+Y \geq 120 \mid X = 50)$ は【キ】ですね。

A： 50 点未満の人の合格率はこの確率より小さいのだから，確かに考えてみてもよいかな？

A： もう少し午前の試験について教えてほしいのだけど，午前の試験が 50 点以上の人の午前の試験の期待値はいくらくらいなのかな？

B： $E[X \mid X \geq 50]$ のことですか？それはおおよそ【ク】ですね。

A： 午前の試験が 50 点以上の人の午後の試験の期待値もわかる？

B： $E[Y \mid X \geq 50]$ のことですか？それは【ケ】ですね。

A： 午後の試験の期待値の方が悪くなるのか。なぜなんだろう？

B： 平均への回帰ですよ。

A： えっ?? とにかく，ありがとう。いろんなことがわかったよ。考える価値はありそうだね。

問 3　A 大学の K 准教授が担当するゼミには 10 名の学生がいる。これらの学生のうち，6 名が学生生活に満足していて，4 名が不満を感じている。これら 10 名の中からランダムに 5 名を選び，第 i 番目 $(i = 1, \ldots, 5)$ の学生の（満足／不満）を表す確率変数

$$X_i = \begin{cases} 1 & （学生生活に満足） \\ 0 & （学生生活に不満） \end{cases}$$

を考える。以下の各問に答えよ。

〔1〕　X_i の期待値 $E[X_i]$ および分散 $V[X_i]$ を求めよ。

〔2〕　X_i, X_j $(i \neq j)$ に対し，積の期待値 $E[X_iX_j]$ と共分散 $Cov[X_i, X_j]$ および相関係数 $R[X_i, X_j]$ を求めよ。

〔3〕　選ばれた 5 名の学生で「満足」と答える人数，すなわち和 $T_5 = \sum_{i=1}^{5} X_i$ の期待値 $E[T_5]$ および分散 $V[T_5]$ を求めよ。

以下では 10 名中，学生生活に満足している学生数を一般に M とする。

K 准教授がゼミ生 10 名からランダムに選んだ 5 名について調査したところ，5 名全員が学生生活に満足していると答えた。すなわち $T_5 = 5$ であった。K 准教授は，ゼミ生のほぼ全員が学生生活に満足しているものと考え，胸をなでおろした。以下の各問に答えよ。

〔4〕　$M = 6$ のときに $T_5 = 5$ となる確率はいくらか。

〔5〕　$T_5 = 5$ が観測されたときに，M の 95%片側信頼区間に含まれる最小の M の値はいくらか。

問 4　メトリックな多次元尺度法 (MDS) は，個体間の非類似度（距離）のデータから，各個体の平面などでの布置を求める手法である。全部で n 個の個体があり，個体 i と個体 j の間の非類似度を $d_{ij}^2 \geq 0$ とする $(i, j = 1, \ldots, n)$。ここでは，$d_{ij}^2 = d_{ji}^2$ $(i > j)$ および $d_{ii}^2 = 0$ $(i = 1, \ldots, n)$ とする。そして，d_{ij}^2 を第 (i, j) 要素とする n 次正方行列（非類似度行列）を D とし，n 次の中心化行列を $Q_n = I_n - \dfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ とする。ただし，I_n は n 次の単位行列，$\mathbf{1}_n$ は成分がすべて 1 の n 次列ベクトルで，上付き添え字 ^{T} は行列もしくはベクトルの転置を表す。このとき，多次元尺度法は以下の (a) 〜 (d) の手順で行われる。

(a)　行列 $P = -\dfrac{1}{2}Q_nDQ_n$ を求める（この変換をヤング・ハウスホルダー変換という）。

(b)　P のスペクトル分解（固有値・固有ベクトル分解）$P = U\Delta U^T$ を求める。ここで，Δ は P の固有値 $(\lambda_1 \geq \cdots \geq \lambda_n)$ を対角要素に持つ対角行列，U は対応する固有ベクトルを列ベクトルとした直交行列である。

(c)　P の正の値を取る固有値を大きい方から k 個選び，それらの正の平方根を対角要素に持つ k 次の対角行列を $\Delta_k^{1/2}$ とし，対応する固有ベクトルを各列に持つ $n \times k$ 行列を U_k として，$\tilde{X} = U_k\Delta_k^{1/2}$ を求める。

(d)　$\tilde{X} = U_k\Delta_k^{1/2}$ の各行ベクトルを k 次元空間（$k = 2$ であれば平面）に布置する。

このとき，以下の各問に答えよ。

〔1〕　具体的に $n = 3$ とし，3 つの個体間の非類似度行列は

$$D = \begin{pmatrix} 0 & 4 & 2 \\ 4 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix} \tag{1}$$

であるとする。中心化行列は

$$Q_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

となる。

〔1-1〕　式 (1) の行列に対し，上記の手順 (a) における行列 P を求めよ。

〔1-2〕　P の固有値は $\lambda_1 = 2$，$\lambda_2 = 2/3$，$\lambda_3 = 0$ であり，λ_1 および λ_2 に対応する固有ベクトルはそれぞれ $\boldsymbol{u}_1 = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$，$\boldsymbol{u}_2 = \dfrac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$ であることが示される。このとき，上記の手順 (c) の行列 $\tilde{X} = U_2\Delta_2^{1/2}$ を求めた上で，$\tilde{X}$ の 3 つの行ベクトルを 2 次元平面に布置せよ。

〔2〕　一般に，k 次元ユークリッド空間 R^k における第 i 個体の座標を表す行ベクトルを $\boldsymbol{x}_i^T = (x_{i1}, \ldots, x_{ik})$ とし，それらを各行に持つ $n \times k$ 行列 $(n > k)$ を X とする。そして，$\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ 間のユークリッド距離の 2 乗 d_{ij}^2 を第 (i, j) 要素に持つ n 次正方行列を D とする。

〔2-1〕　D を XX^T および $\mathbf{1}_n$ を用いて表現せよ。

〔2-2〕　D をヤング・ハウスホルダー変換した $P = -\dfrac{1}{2} Q_n D Q_n$ を Q_n と X を用いて表せ。

〔3〕　サンプルサイズ n，変量数 p の分散共分散行列に基づく主成分分析において，k 個の主成分得点からなる $n \times k$ 行列を Z としたとき，Z と上記の手順 (c) で求めた $\tilde{X}$ との関係を示せ。

問 5　ある IT 会社では，新卒採用の各応募者に対して 2 段階の試験を実施している。第 1 段階は数学と英語の 2 科目からなる一般教養試験で，第 2 段階は採用に直結した情報の試験である。同社の採用担当の A さんは，第 1 段階の一般教養試験の点数から第 2 段階の情報の試験の点数を予測しようとした。以下では，x_1 を数学の点数，x_2 を英語の点数，y を情報の点数とし，それらの点数はそれぞれ平均 0，分散 1 に標準化されているものとする。また，y と x_1，y と x_2 および x_1 と x_2 間の相関係数をそれぞれ r_{y1}，r_{y2}，r_{12} とする。

A さんが調べたデータの各試験科目間の相関行列は

$$R = \begin{pmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0 \\ 0.6 & 0 & 1 \end{pmatrix} \tag{1}$$

であった。A さんは，英語 (x_2) と情報 (y) 間の相関係数 r_{y2} は 0 であったので，英語は情報の予測に影響を与えないと考え，数学 (x_1) のみを説明変数とする単回帰式を求めたところ $y = 0.6x_1$ となり，このときの決定係数は $R_1^2 = 0.6^2 = 0.36$ であった。しかし同僚の B さんから，x_2 もモデルに入れてみたらとのアドバイスがあり，改めて重回帰式を計算したところ

$$y = 0.8x_1 - 0.4x_2 \tag{2}$$

を得て，決定係数 R_2^2 も大きくなった。A さんは，y との相関が 0 の x_2 をモデルに加えることで x_1 の係数が 0.2 増加して予測力が上がり，モデルの当てはまりがよくなったのが不思議であった。

この分析結果を上司の C さんに報告したところ，C さんから，式 (2) の重回帰式の英語の係数が負であるが，英語力があるほど情報の点数が低いという結果は，IT 業界では英語は不可欠であるので問題であり，A さんの計算が間違っているか，あるいは第 2 段階の情報の試験問題は不適切だったのではないかとの意見があった。

以下の各問に答えよ。なお，一般の相関行列を考察する際は $0 \leq r_{12} < 1$ および $r_{y1} > 0$ とする。

〔1〕　説明変数 x_1，x_2 から y を予測する重回帰式 $y = b_1x_1 + b_2x_2$ の係数 b_1，b_2 の最小二乗法による推定値は，正規方程式

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix}$$

の解として与えられる。正規方程式から b_1 と b_2 を求める計算式を示し，式 (1) の相関行列の数値を代入することで，式 (2) の A さんの計算結果を確認せよ。

〔2〕　A さんの計算では，重回帰式の x_1 の係数 b_1 は単回帰式の x_1 の係数 r_{y1} よりも大きくなった。上問〔1〕の b_1 の計算式を用いて，$b_1 > r_{y1}$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔3〕　A さんの計算では，$r_{y2} = 0$ であったが重回帰式での x_2 の係数 b_2 の符号が負となった。一般に，$r_{y2} \geq 0$ であるが $b_2 < 0$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔4〕　式 (2) の重回帰式による y の予測値を $\hat{y} = 0.8x_1 - 0.4x_2$ としたとき，$\hat{y}$ の分散および y と $\hat{y}$ の共分散をそれぞれ求めよ。

〔5〕　式 (2) の重回帰式の決定係数 R_2^2 はいくらか。

社会科学

問 1　ある母集団における変量 x の母平均は μ である。その母集団は H 個の互いに異なる層（グループ）に分けられていて，第 h 層の比率を w_h $(w_h \geq 0, w_1+\cdots+w_H=1)$ とし，その層での変量 x の標準偏差を σ_h とする $(h=1,\ldots,H)$。ここでは，すべての層は十分大きく，それぞれ無限母集団とみなすことができるとする。

母集団の第 h 層から，他の層とは独立に大きさ n_h の標本を非復元無作為抽出し，その層における変量 x の標本平均を $\bar{X}_h$ とする。表 1 は層化無作為抽出における各層の情報をまとめたものである。

表 1：層化無作為抽出の概要

	層番号	1	2	$\cdots$	h	$\cdots$	H
母集団	比率	w_1	w_2	$\cdots$	w_h	$\cdots$	w_H
	標準偏差	σ_1	σ_2	$\cdots$	σ_h	$\cdots$	σ_H
		$\downarrow$	$\downarrow$		$\downarrow$		$\downarrow$
標本	大きさ	n_1	n_2	$\cdots$	n_h	$\cdots$	n_H
	平均	$\bar{X}_1$	$\bar{X}_2$	$\cdots$	$\bar{X}_h$	$\cdots$	$\bar{X}_H$

このような抽出が行われたとき，母平均 μ の不偏推定量として

$$\bar{X}_{st} = \sum_{h=1}^{H} w_h \bar{X}_h \tag{1}$$

が用いられ，推定量の分散は

$$V[\bar{X}_{st}] = \sum_{h=1}^{H} w_h^2 \cdot \frac{1}{n_h}\sigma_h^2 \tag{2}$$

となる。全体の標本の大きさを $n=\sum_{h=1}^{H} n_h$ とするとき，式 (2) を最小にする n_h $(h=1,\ldots,H)$ は

$$n_h = \frac{w_h\sigma_h}{\sum_{k=1}^{H} w_k\sigma_k}\cdot n \tag{3}$$

によって得られ，このときの配分法はネイマン配分法（最適配分法）と呼ばれる。以下の各問に答えよ。

〔1〕　ネイマン配分法を用いた場合の推定量の分散は

$$V[\bar{X}_{st}] = \frac{1}{n}\left(\sum_{h=1}^{H} w_h\sigma_h\right)^2$$

となることを示せ。

〔2〕　A 市にある大規模な 4 つの大学に所属する全学生を母集団とし，それぞれの大学を母集団の第 1 層～第 4 層とする。学生の 1 日当たりの睡眠時間を知るため，合計 300 人の学生をネイマン配分法により層化無作為抽出して睡眠時間を調査し，そのデータを基に 4 大学の全学生の母平均を推定する。

表 2 は母集団における 4 大学の学生の比率と変量の標準偏差を示している。ここで比率とは，その大学の学生数を 4 大学の学生数の合計で割った値である。また，4 大学の学生数はそれぞれ十分大きくて無限母集団とみなせるとし，標準偏差を含めて表 2 の値はすべて既知とする。

表 2：母集団における各層の情報

層番号	1	2	3	4
比率	0.1	0.3	0.4	0.2
標準偏差	1.6	1.8	1.1	1.8

〔2-1〕　標本を抽出したところ，各層の標本平均は $\bar{X}_1 = 6.6$，$\bar{X}_2 = 6.8$，$\bar{X}_3 = 7.2$，$\bar{X}_4 = 6.2$ であった。式 (1) を用いて母平均 μ を推定せよ。

〔2-2〕　ネイマン配分法により各大学から得る標本の大きさをそれぞれ求めよ。

〔2-3〕　推定量の分散 $V[\bar{X}_{st}]$ を求めよ。

〔3〕　母平均を推定する時点において，母集団の各層の標準偏差が既知であることは少ない。ここでは上問〔2〕と同じ設定において，母集団の 4 大学の学生の比率 w_h は既知であるが，標準偏差 σ_h は未知とする。そのため，ネイマン配分法を用いて層化無作為抽出する調査（本調査と呼ぶ）の前に，標準偏差のおおよその値を見積もるための予備調査を行う。予備調査では合計 100 人を比率 w_h に基づく比例配分法で抽出し，本調査では合計 200 人をネイマン配分法で抽出する。

表 3 は，予備調査で抽出された標本における各層の標本平均と標準偏差である。ここでは，予備調査で得られた各層の標準偏差を母集団の真の標準偏差とみなす。

表 3：予備調査で得られた標本における各層の情報

層番号	1	2	3	4
平均	6.4	6.9	7.1	6.9
標準偏差	1.2	2.0	1.2	2.0

〔3-1〕　本調査で抽出される標本から計算される推定量 $\bar{X}_{st}^{M}$ の分散 $V[\bar{X}_{st}^{M}]$ を求めよ。

〔3-2〕　本調査で得られる情報を用いず，予備調査で得られた情報だけを用いて母平均 μ を式 (1) で推定し，その値を $\bar{X}_{st}^{P}$ としたとき，その分散 $V[\bar{X}_{st}^{P}]$ を求めよ。

〔3-3〕　上問〔3-1〕および〔3-2〕のそれぞれの推定量に対し，$\hat{\mu} = \alpha_1 \bar{X}_{st}^{M} + \alpha_2 \bar{X}_{st}^{P}$ として μ を不偏推定する。推定量 $\bar{X}_{st}^{M}$ と $\bar{X}_{st}^{P}$ が独立であるとき，推定量 $\hat{\mu}$ の分散 $V[\hat{\mu}]$ を最小にする α_1 と α_2 を求めよ。また，そのときの分散 $V[\hat{\mu}]$ はいくらか。

問 2　ポイントの多寡を競うゲームがあり，表 1 は，ゲームに参加した 20 人について，そのゲームで得たポイント別の人数を集計したものである。以下の各問に答えよ。

表 1：20 人のゲームのポイントの集計結果

ポイント	1	2	3	4	5	6	7	8
人数	1	1	3	3	4	2	4	2

〔1〕　表 1 の集計結果から，下のようなポイントごとの累積人数比率と累積ポイント比率の表を解答用紙に作成せよ。

ポイント	1	2	3	4	5	6	7	8
累積人数比率								1.00
累積ポイント比率								1.00

その表を基に，横軸に累積人数比率，縦軸に累積ポイント比率をとって，完全平等線（45° 線）とともにポイントに関するローレンツ曲線を描け。

〔2〕　表 1 のデータについて，ポイントを 2 つずつまとめて新たに表 2 を作成した。

表 2：ポイントを 2 つずつまとめた結果

ポイント（階級）	1〜2	3〜4	5〜6	7〜8
人数	2	6	6	6
ポイント（合計）	3	21	32	44

〔2-1〕　表2のような4つの階級のデータから作成された累積人数比率と累積ポイント比率を表3のように文字を使って表すとき，ジニ係数をこれらの文字の多項式として簡単に表せ。

表3：ポイントを2つずつまとめた結果

ポイント（階級）	1〜2	3〜4	5〜6	7〜8
累積人数比率	x_1	x_2	x_3	1
累積ポイント比率	y_1	y_2	y_3	1

〔2-2〕　表2のデータからジニ係数を求めよ。

〔2-3〕　人数の分布が表2と同じで，ポイント（合計）の分布が表4のような5つのデータを考える。表2のデータは，表4の「データ3」に対応する。上問〔2-1〕の結果を用いて，5つのデータのうちでジニ係数が最大のデータがどれであるかを述べ，そのジニ係数を求めよ。

表4：5つのデータでそれぞれポイントを2つずつまとめた結果

ポイント（階級）	1〜2	3〜4	5〜6	7〜8
データ1のポイント（合計）	3	21	30	46
データ2のポイント（合計）	3	21	31	45
データ3のポイント（合計）	3	21	32	44
データ4のポイント（合計）	3	21	33	43
データ5のポイント（合計）	3	21	34	42

問 3　確率過程 $\{x_t\}$ は式 (1) の 2 次の自己回帰 (AR(2)) モデルに従っているとする。

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t \tag{1}$$

ここで，誤差項 $\{\varepsilon_t\}$ は以下の 3 条件を満たす確率変数の列とする。

- 任意の整数 t に対して期待値は $E[\varepsilon_t] = 0$ で分散は $V[\varepsilon_t] = \sigma^2 < \infty$
- $E[\varepsilon_t \varepsilon_s] = 0$（$t$，$s$ は整数；$t \neq s$）
- $E[x_{t-s}\varepsilon_t] = 0$（$t$ は整数；$s = 1, 2, 3, \ldots$）

さらに，モデル (1) は弱定常であると仮定する。また k 時点差の自己相関を $\rho(k)$ とし，k 時点差の偏自己相関（つまり x_t と x_{t-k} の偏自己相関）を ϕ_{kk} とする。以下の各問に答えよ。

なお，偏自己相関と自己相関の間には

$$\phi_{11} = \rho(1)$$

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \end{pmatrix}$$

の関係があることを用いてよい。ここで，ϕ_{kj} $(k > j)$ は x_t と x_{t-j} の関係を表す係数で，偏自己相関 ϕ_{kk} の導出のために用いられるものである。

〔1〕　パラメータの組合せが $\phi_1 = 1$，$\phi_2 = -0.4$ のとき確率過程 $\{x_t\}$ は弱定常になる。その理由を示せ。

〔2〕　モデル (1) における 1 時点差，2 時点差，3 時点差の自己相関 $\rho(1)$，$\rho(2)$，$\rho(3)$ を ϕ_1 および ϕ_2 を用いて表せ。

〔3〕　モデル (1) における 1 時点差，2 時点差，3 時点差の偏自己相関 ϕ_{11}，ϕ_{22}，ϕ_{33} を ϕ_1 および ϕ_2 を用いて表せ。

〔4〕　偏自己相関は，確率過程 $\{x_t\}$ のどのような特徴を表しているか。自己相関との違いに基づいて述べよ。

〔5〕　時系列 x_t $(t = 1, \ldots, T)$ の観測値が得られたとする。標本平均を $\bar{x} = (1/T)\sum_{t=1}^{T} x_t$ とし，標本分散，標本自己共分散，標本自己相関を以下のように定める。

$$\hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

$$\hat{\gamma}(s) = \frac{1}{T}\sum_{t=s+1}^{T}(x_t - \bar{x})(x_{t-s} - \bar{x}), \quad \hat{\rho}(s) = \frac{\hat{\gamma}(s)}{\hat{\gamma}(0)} \quad (s = 0, 1, 2, \ldots)$$

このとき，AR(2) モデルに基づいた ϕ_1，ϕ_2 のユール・ウォーカー推定量 $\hat{\phi}_1$，$\hat{\phi}_2$ を求めよ。

問 4　ある企業は社員の英語能力の向上に力を入れていて，そのために英語講習の強化プログラムと普通プログラムの 2 種類を用意している。社員は，年度初めの 4 月に英語のテストを受け，その結果に応じて自らの判断で強化プログラムか普通プログラムのいずれかに登録し，年度終りに講習の成果を見るため年度初めと同レベルのテストを受けた。プログラムの種類を表す変数を c とし，強化プログラムを $c = 1$，普通プログラムを $c = 0$ とする。それぞれのプログラムでの講習前のテスト得点を X_c，講習後のテスト得点を Y_c としたとき，テスト得点の組 $\begin{pmatrix} X_c \\ Y_c \end{pmatrix}$ は 2 変量正規分布 $N\left(\begin{pmatrix} \mu_{X_c} \\ \mu_{Y_c} \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_c} & \sigma_{XY_c} \\ \sigma_{XY_c} & \sigma^2_{Y_c} \end{pmatrix}\right)$ $(c = 1, 0)$ に従うと仮定する。また，テスト得点の差 $D_c = Y_c - X_c$ を変化量と呼ぶ。

強化プログラムと普通プログラムの両方からそれぞれ 20 名ずつの社員をランダムに選んでテストの点数を調べたところ，表 1 のような結果を得た。ただし，標準偏差はそれぞれ偏差平方和の除数を 19 とした不偏分散の正の平方根である。講習前のテストの点数が低かった社員の多くは強化プログラムを選択していたようである。以下の各問に答えよ。

表 1：社員のテスト得点の要約

強化プログラム	講習前 (X_1)	講習後 (Y_1)
平均	50	52
標準偏差	10	8
相関係数	0.5	

普通プログラム	講習前 (X_0)	講習後 (Y_0)
平均	60	61
標準偏差	10	8
相関係数	0.5	

〔1〕　表 1 から，強化プログラムと普通プログラムのそれぞれに対し，テスト得点の変化量の平均と分散を求めよ。

〔2〕　テスト得点の変化量について，強化プログラムの母平均 $\delta_1 = \mu_{Y_1} - \mu_{X_1}$ と普通プログラムの母平均 $\delta_0 = \mu_{Y_0} - \mu_{X_0}$ が等しいかどうかの有意水準 5% の t 検定を行い，その結果を述べよ。

〔3〕　強化プログラムと普通プログラムのそれぞれに対し，講習前のテスト得点 (x) から講習後のテスト得点 (y) を予測する単回帰式 $y = a_c + b_c x$ を求めよ。

〔4〕　強化プログラムと普通プログラムのそれぞれにおいて，上問〔3〕の単回帰式により，講習前のテスト得点が $x = 55$ の社員の講習後のテスト得点 Y_c の条件付き期待値 $E[Y_c \mid x = 55]$ の各推定値 m_1 および m_0 を求めよ。また，m_c の条件付き分散の推定値は $c = 1$ および $c = 0$ の両方とも同じ $\hat{V}[m_c \mid x = 55] = 3.03$ であるとして，強化プログラムと普通プログラム間に条件付き期待値 $E[Y_c \mid x = 55]$ の差があるかどうかの有意水準 5% の検定を行い，その結果を述べよ。

〔5〕　上問〔2〕と〔4〕の結果を踏まえ，強化プログラムと普通プログラムにおける講習の効果について述べよ。

問 5　ある IT 会社では，新卒採用の各応募者に対して 2 段階の試験を実施している。第 1 段階は数学と英語の 2 科目からなる一般教養試験で，第 2 段階は採用に直結した情報の試験である。同社の採用担当の A さんは，第 1 段階の一般教養試験の点数から第 2 段階の情報の試験の点数を予測しようとした。以下では，x_1 を数学の点数，x_2 を英語の点数，y を情報の点数とし，それらの点数はそれぞれ平均 0，分散 1 に標準化されているものとする。また，y と x_1，y と x_2 および x_1 と x_2 間の相関係数をそれぞれ r_{y1}，r_{y2}，r_{12} とする。

A さんが調べたデータの各試験科目間の相関行列は

$$R = \begin{pmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0 \\ 0.6 & 0 & 1 \end{pmatrix} \tag{1}$$

であった。A さんは，英語 (x_2) と情報 (y) 間の相関係数 r_{y2} は 0 であったので，英語は情報の予測に影響を与えないと考え，数学 (x_1) のみを説明変数とする単回帰式を求めたところ $y = 0.6x_1$ となり，このときの決定係数は $R_1^2 = 0.6^2 = 0.36$ であった。しかし同僚の B さんから，x_2 もモデルに入れてみたらとのアドバイスがあり，改めて重回帰式を計算したところ

$$y = 0.8x_1 - 0.4x_2 \tag{2}$$

を得て，決定係数 R_2^2 も大きくなった。A さんは，y との相関が 0 の x_2 をモデルに加えることで x_1 の係数が 0.2 増加して予測力が上がり，モデルの当てはまりがよくなったのが不思議であった。

この分析結果を上司の C さんに報告したところ，C さんから，式 (2) の重回帰式の英語の係数が負であるが，英語力があるほど情報の点数が低いという結果は，IT 業界では英語は不可欠であるので問題であり，A さんの計算が間違っているか，あるいは第 2 段階の情報の試験問題は不適切だったのではないかとの意見があった。

以下の各問に答えよ。なお，一般の相関行列を考察する際は $0 \leq r_{12} < 1$ および $r_{y1} > 0$ とする。

〔1〕　説明変数 x_1，x_2 から y を予測する重回帰式 $y = b_1x_1 + b_2x_2$ の係数 b_1，b_2 の最小二乗法による推定値は，正規方程式

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix}$$

の解として与えられる。正規方程式から b_1 と b_2 を求める計算式を示し，式 (1) の相関行列の数値を代入することで，式 (2) の A さんの計算結果を確認せよ。

〔2〕　A さんの計算では，重回帰式の x_1 の係数 b_1 は単回帰式の x_1 の係数 r_{y1} よりも大きくなった。上問〔1〕の b_1 の計算式を用いて，$b_1 > r_{y1}$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔3〕　A さんの計算では，$r_{y2} = 0$ であったが重回帰式での x_2 の係数 b_2 の符号が負となった。一般に，$r_{y2} \geq 0$ であるが $b_2 < 0$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔4〕　式 (2) の重回帰式による y の予測値を $\hat{y} = 0.8x_1 - 0.4x_2$ としたとき，$\hat{y}$ の分散および y と $\hat{y}$ の共分散をそれぞれ求めよ。

〔5〕　式 (2) の重回帰式の決定係数 R_2^2 はいくらか。

理工学

問 1　あるセラミック焼成工程では，焼成後のセラミック強度を改善するために，A：焼成温度（2 水準：A_1，A_2），B：原料 B の配合量（2 水準：B_1，B_2），C：原料 C の配合量（3 水準：C_1，C_2，C_3）の 3 因子を取り上げて実験を行い，最も強度の高くなる焼成温度と原料の配合を求めたいと考えた。3 因子のうち，焼成温度は水準設定が困難であるので，A を 1 次因子とし，B，C を 2 次因子として分割実験を行うこととした。また，A，B，C の異なる水準の各組合せについて 3 回ずつ反復し，各ブロックサイズが 12 である 3 個のブロックで実験することとした。なお簡単のため，このブロック因子は固定効果とする。この実験計画につき，以下の各問に答えよ。

〔1〕　この分割法で得られる 36 回の実験について，その実験順序はどのように定めればよいか説明せよ。

〔2〕　水準 (A_i, B_j, C_k) の反復 R_l における応答値を y_{ijkl} とする $(i = 1, 2; j = 1, 2; k = 1, 2, 3; l = 1, 2, 3)$。そして，定数項を μ とし，因子 A，B，C の主効果を順に α，β，γ，A×B の交互作用を $(\alpha\beta)$，A×C の交互作用を $(\alpha\gamma)$，B×C の交互作用を $(\beta\gamma)$，A×B×C の 3 因子交互作用を $(\alpha\beta\gamma)$，反復の効果を ρ，1 次誤差を $\varepsilon_{(1)}$，2 次誤差を $\varepsilon_{(2)}$ とする。このとき，適切に添え字をつけて y_{ijkl} に対する構造式を記述せよ。ただし，定数項以外の主効果ならびに各交互作用項の和は 0 とし，誤差分布に関する記載は省略してよい。

〔3〕　この実験において，反復を行わなかった場合，上問〔2〕の構造式はどのようになるか説明し，構造式を記述せよ。

〔4〕　1 次誤差と 2 次誤差の分散をそれぞれ $\sigma^2_{(1)}$，$\sigma^2_{(2)}$ とする。このとき，1 次誤差と A の主効果の平方和はそれぞれ

$$S_{E(1)} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{3}\left(\bar{y}_{i\cdot\cdot l} - \bar{y}_{i\cdots} - \bar{y}_{\cdots l} + \bar{y}\right)^2 = 6\sum_{i=1}^{2}\sum_{l=1}^{3}\left(\bar{y}_{i\cdot\cdot l} - \bar{y}_{i\cdots} - \bar{y}_{\cdots l} + \bar{y}\right)^2$$

$$S_A = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{3}\left(\bar{y}_{i\cdots} - \bar{y}\right)^2 = 18\sum_{i=1}^{2}\left(\bar{y}_{i\cdots} - \bar{y}\right)^2$$

となる。これらの平方和の期待値 $E[S_{E(1)}]$，$E[S_A]$ を計算せよ。ただし，下付き添え字におけるドット $(\cdot)$ はその添え字に関する平均を表す記号である。

〔5〕　実際の分割実験の結果を繰り返しのない 4 因子完全無作為化実験とみなして，それぞれの平方和と自由度を求めたところ表 1 のようになった。表 1 の数値をもとに表 2 の分散分析表を作成するとき，空欄（ア）～（キ）に当てはまる数値を答えよ。また，要因 A のセラミック強度への効果に関する有意水準 5%の検定の有意性について論ぜよ。

表1：セラミック焼成工程の実験結果

要因	平方和	自由度
R	28.741	2
A	11.674	1
B	40.704	1
C	26.360	2
A×R	13.133	2
B×R	0.022	2
C×R	2.884	4
A×B	7.489	1
A×C	2.809	2
B×C	19.014	2
A×B×R	2.843	2
A×C×R	4.462	4
B×C×R	6.845	4
A×B×C	1.780	2
A×B×C×R	2.772	4
計	171.532	35

表2：分散分析表

要因	平方和	自由度	平均平方	F 値
R				
A	（ア）	（イ）	（ウ）	（エ）
1次誤差	（オ）	（カ）	（キ）	
B				
C				
A×B				
A×C				
B×C				
A×B×C				
2次誤差				
計	171.532	35		

問 2　ある 2 種類の製品について耐久性試験を行い，製品が故障するまでの時間を調べたところ，表 1 のデータが得られた。ただし x は製品の種類を表すダミー変数であり，T は故障時間（日）を表す。

表 1：耐久性試験の結果

実験番号 i	1	2	3	4
種類 x	0	0	1	1
故障時間 T	35	16	9	12

このデータに対して統計モデルを当てはめる。故障時間の累積分布関数を $F(t)$，ハザード関数を $h(t)$ と置く。$F(t)$ と $h(t)$ の間には

$$F(t) = 1 - \exp\left(-\int_0^t h(s)ds\right)$$

という関係がある。表 1 における実験番号 i のデータを $(x_i; T_i)$ と記す。このとき，以下の各問に答えよ。

〔1〕　ハザード関数が t によらず $h(t) = \exp(\alpha + \beta x)$ であるとする。ただし，α，β はパラメータである。

〔1-1〕　累積分布関数 $F(t)$ を求めよ。

〔1-2〕　表 1 のデータに基づく尤度関数を求めよ。

〔1-3〕　$\theta_0 = \exp(\alpha)$，$\theta_1 = \exp(\alpha + \beta)$ と置き，θ_0，θ_1 の最尤推定値を求めよ。その結果を利用して α，β の最尤推定値を求めよ。

〔2〕　ハザード関数を

$$h(t) = \gamma t^{\gamma-1} \exp(\alpha + \beta x)$$

と仮定する。ただし，α，β，γ はパラメータである。

〔2-1〕　表 1 のデータに基づく尤度関数を求めよ。

〔2-2〕　対数尤度関数の α，β，γ に関するヘッセ行列（2 階導関数を並べてできる行列）が負定値（固有値がすべて負）であることを示せ。

問 3　確率過程 $\{x_t\}$ は式 (1) の 2 次の自己回帰 (AR(2)) モデルに従っているとする。

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t \tag{1}$$

ここで，誤差項 $\{\varepsilon_t\}$ は以下の 3 条件を満たす確率変数の列とする。

- 任意の整数 t に対して期待値は $E[\varepsilon_t] = 0$ で分散は $V[\varepsilon_t] = \sigma^2 < \infty$
- $E[\varepsilon_t \varepsilon_s] = 0$（$t$，$s$ は整数；$t \neq s$）
- $E[x_{t-s}\varepsilon_t] = 0$（$t$ は整数；$s = 1, 2, 3, \ldots$）

さらに，モデル (1) は弱定常であると仮定する。また k 時点差の自己相関を $\rho(k)$ とし，k 時点差の偏自己相関（つまり x_t と x_{t-k} の偏自己相関）を ϕ_{kk} とする。以下の各問に答えよ。

なお，偏自己相関と自己相関の間には

$$\phi_{11} = \rho(1)$$

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \end{pmatrix}$$

の関係があることを用いてよい。ここで，ϕ_{kj} $(k > j)$ は x_t と x_{t-j} の関係を表す係数で，偏自己相関 ϕ_{kk} の導出のために用いられるものである。

〔1〕　パラメータの組合せが $\phi_1 = 1$，$\phi_2 = -0.4$ のとき確率過程 $\{x_t\}$ は弱定常になる。その理由を示せ。

〔2〕　モデル (1) における 1 時点差，2 時点差，3 時点差の自己相関 $\rho(1)$，$\rho(2)$，$\rho(3)$ を ϕ_1 および ϕ_2 を用いて表せ。

〔3〕　モデル (1) における 1 時点差，2 時点差，3 時点差の偏自己相関 ϕ_{11}，ϕ_{22}，ϕ_{33} を ϕ_1 および ϕ_2 を用いて表せ。

〔4〕　偏自己相関は，確率過程 $\{x_t\}$ のどのような特徴を表しているか。自己相関との違いに基づいて述べよ。

〔5〕　時系列 x_t $(t = 1, \ldots, T)$ の観測値が得られたとする。標本平均を $\bar{x} = (1/T)\sum_{t=1}^{T} x_t$ とし，標本分散，標本自己共分散，標本自己相関を以下のように定める。

$$\hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

$$\hat{\gamma}(s) = \frac{1}{T}\sum_{t=s+1}^{T}(x_t - \bar{x})(x_{t-s} - \bar{x}), \quad \hat{\rho}(s) = \frac{\hat{\gamma}(s)}{\hat{\gamma}(0)} \quad (s = 0, 1, 2, \ldots)$$

このとき，AR(2) モデルに基づいた ϕ_1，ϕ_2 のユール・ウォーカー推定量 $\hat{\phi}_1$，$\hat{\phi}_2$ を求めよ。

問 4　ある医療機器メーカーは，脳画像撮影のための MRI (magnetic resonance imaging) 検査装置を多くの医療施設に納入していて，メーカーの品質保証部では，定期的にいくつかの医療施設を選定し，施設ごとに収集した検査データに基づいて機器の動作モニタリングを行っている。品質保証部の A さんは，検査データは施設ごとに平均が多少異なることに気付いていて，各施設からのデータを調整して施設間差を無くした上で，検査データ全体の分布を求めようとしている。

A さんは，医療施設の中から無作為に m 箇所を抽出し，各施設から MRI 検査データをそれぞれ n 人ずつ入手した。ただし，施設ごとのデータは少数個（例えば 3 人分ずつ）であった。そしてデータのある特徴量につき，第 j 施設の第 i 測定値を表す確率変数を Y_{ij} として，モデル

$$Y_{ij} = \alpha + \gamma_j + \beta x_{ij} + \varepsilon_{ij} \quad (i = 1\ldots, n; j = 1, \ldots, m) \tag{1}$$

を想定した。ここで，x_{ij} は測定値に影響を及ぼす患者の年齢，α と β は施設によらない未知の定数パラメータ，γ_j は第 j 施設の測定値を特徴付ける未知パラメータ，ε_{ij} は互いに独立に正規分布 $N(0, \sigma^2)$ に従う確率変数である。また，Y_{ij} の観測値を y_{ij} とし，それらから求めた全平均および施設ごとの平均を

$$\bar{y} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} y_{ij}, \quad \bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij} \quad (j = 1, \ldots, m)$$

とする。以下の各問に答えよ。

〔1〕　仮に施設間差がないとする。すなわち，式 (1) で $\gamma_j = 0$ $(j = 1, \ldots, m)$ の場合である。このときの α と β の最小二乗推定値をそれぞれ a，b とし，Y_{ij} の予測値を $\hat{y}_{ij} = a + bx_{ij}$，残差を $r_{ij} = y_{ij} - \hat{y}_{ij}$ とする。

〔1-1〕　最小二乗推定値 a，b の式をそれぞれ求めよ。

〔1-2〕　予測値の平均は観測値の平均に等しく，予測値と残差との相関は 0 であることを示せ。

〔1-3〕　観測値，予測値，残差の偏差平方和をそれぞれ

$$S_T = \sum_{j=1}^{m}\sum_{i=1}^{n} (y_{ij} - \bar{y})^2, \quad S_M = \sum_{j=1}^{m}\sum_{i=1}^{n} (\hat{y}_{ij} - \bar{y})^2, \quad S_R = \sum_{j=1}^{m}\sum_{i=1}^{n} (y_{ij} - \hat{y}_{ij})^2$$

とするとき，$S_T = S_M + S_R$ となることを示せ。

〔2〕　式 (1) で，γ_j は施設ごとに異なる定数であり，$\sum_{j=1}^{m}\gamma_j = 0$ とする。

〔2-1〕　α，β，γ_j の最小二乗推定値 a，b，g_j をそれぞれ求めよ。

〔2-2〕　A さんは，施設間差を調整した調整値として

$$y_{ij}^{(A)} = y_{ij} - g_j$$

を用いた。モデル (1) の下で，$y_{ij}^{(A)}$ を確率変数とみた場合，その期待値はいくらか。

〔3〕　式 (1) で，γ_j は互いに独立かつ ε_{ij} とも独立に正規分布 $N(0, \tau^2)$ に従う確率変数とする。

〔3-1〕　Y_{ij} の従う分布は何か。その分布の期待値と分散と共に答えよ。

〔3-2〕　A さんの同僚の B さんは，τ^2 と σ^2 の何らかの推定値を $\tilde{\tau}^2$ および $\tilde{\sigma}^2$ として（これらは過去のデータから精度良く推定されていると仮定），第 j 施設の母平均を重み付き平均

$$\tilde{y}_j = \frac{\tilde{\sigma}^2/n}{\tilde{\tau}^2 + (\tilde{\sigma}^2/n)}\bar{y} + \frac{\tilde{\tau}^2}{\tilde{\tau}^2 + (\tilde{\sigma}^2/n)}\bar{y}_j$$

で推定し，$\tilde{\gamma}_j = \tilde{y}_j - \bar{y}$ として，調整値を

$$y_{ij}^{(B)} = y_{ij} - \tilde{\gamma}_j$$

としたらどうかと提案した。B さんの提案は妥当だろうか。その妥当性を判断し，その理由を述べよ。

問 5　ある IT 会社では，新卒採用の各応募者に対して 2 段階の試験を実施している。第 1 段階は数学と英語の 2 科目からなる一般教養試験で，第 2 段階は採用に直結した情報の試験である。同社の採用担当の A さんは，第 1 段階の一般教養試験の点数から第 2 段階の情報の試験の点数を予測しようとした。以下では，x_1 を数学の点数，x_2 を英語の点数，y を情報の点数とし，それらの点数はそれぞれ平均 0，分散 1 に標準化されているものとする。また，y と x_1，y と x_2 および x_1 と x_2 間の相関係数をそれぞれ r_{y1}，r_{y2}，r_{12} とする。

A さんが調べたデータの各試験科目間の相関行列は

$$R = \begin{pmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0 \\ 0.6 & 0 & 1 \end{pmatrix} \tag{1}$$

であった。A さんは，英語 (x_2) と情報 (y) 間の相関係数 r_{y2} は 0 であったので，英語は情報の予測に影響を与えないと考え，数学 (x_1) のみを説明変数とする単回帰式を求めたところ $y = 0.6x_1$ となり，このときの決定係数は $R_1^2 = 0.6^2 = 0.36$ であった。しかし同僚の B さんから，x_2 もモデルに入れてみたらとのアドバイスがあり，改めて重回帰式を計算したところ

$$y = 0.8x_1 - 0.4x_2 \tag{2}$$

を得て，決定係数 R_2^2 も大きくなった。A さんは，y との相関が 0 の x_2 をモデルに加えることで x_1 の係数が 0.2 増加して予測力が上がり，モデルの当てはまりがよくなったのが不思議であった。

この分析結果を上司の C さんに報告したところ，C さんから，式 (2) の重回帰式の英語の係数が負であるが，英語力があるほど情報の点数が低いという結果は，IT 業界では英語は不可欠であるので問題であり，A さんの計算が間違っているか，あるいは第 2 段階の情報の試験問題は不適切だったのではないかとの意見があった。

以下の各問に答えよ。なお，一般の相関行列を考察する際は $0 \leq r_{12} < 1$ および $r_{y1} > 0$ とする。

〔1〕　説明変数 x_1，x_2 から y を予測する重回帰式 $y = b_1x_1 + b_2x_2$ の係数 b_1，b_2 の最小二乗法による推定値は，正規方程式

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix}$$

の解として与えられる。正規方程式から b_1 と b_2 を求める計算式を示し，式 (1) の相関行列の数値を代入することで，式 (2) の A さんの計算結果を確認せよ。

〔2〕　A さんの計算では，重回帰式の x_1 の係数 b_1 は単回帰式の x_1 の係数 r_{y1} よりも大きくなった。上問〔1〕の b_1 の計算式を用いて，$b_1 > r_{y1}$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔3〕　A さんの計算では，$r_{y2} = 0$ であったが重回帰式での x_2 の係数 b_2 の符号が負となった。一般に，$r_{y2} \geq 0$ であるが $b_2 < 0$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔4〕　式 (2) の重回帰式による y の予測値を $\hat{y} = 0.8x_1 - 0.4x_2$ としたとき，$\hat{y}$ の分散および y と $\hat{y}$ の共分散をそれぞれ求めよ。

〔5〕　式 (2) の重回帰式の決定係数 R_2^2 はいくらか。

医薬生物学

医薬生物学

問 1　以下の各問に答えよ。

〔1〕　疾患 D の罹患者 20 人と非罹患者 20 人に対して，バイオマーカー B の値を測定した。B の値がカットオフ値 c 以上の場合に陽性（罹患あり）と判定し，c 未満の場合に陰性（罹患なし）と判定する。4 種類の c の値 $(0, 10, 20, 30)$ について，罹患者集団と非罹患者集団の陽性数，陰性数，感度，特異度，陽性的中率，陰性的中率を求めた結果の一部を表 1 に示す。

表 1：罹患者集団と非罹患者集団の判定結果

カットオフ値	罹患者集団		非罹患者集団				陽性	陰性
c	陽性	陰性	陽性	陰性	感度	特異度	的中率	的中率
0					1	0	0.5	-
10					0.9	0.4	0.6	0.8
20	12	8	3	17	(a)	(b)	(c)	(d)
30					0	1	-	0.5

〔1-1〕　表 1 の (a) から (d) の値を求めよ。

〔1-2〕　表 1 に示される値から ROC 曲線を描き，その曲線下面積を求めよ。

〔2〕　独立な N 人の被験者から観測された対応のある 2 値データを (X_i, Y_i) $(i = 1, \ldots, N)$ とする。X_i と Y_i は 1 または 0 の値を取り，周辺確率をそれぞれ $P(X_i = 1) = \pi_X$，$P(X_i = 0) = 1 - \pi_X$，$P(Y_i = 1) = \pi_Y$，$P(Y_i = 0) = 1 - \pi_Y$ とし，X_i と Y_i の同時確率を $P(X_i = j, Y_i = k) = \pi_{jk}$ $(j, k = 0, 1)$ とする。また，$Z_i = X_i - Y_i$ と置く。

〔2-1〕　Z_i の期待値 $E[Z_i]$ と分散 $V[Z_i]$ をそれぞれ π_{jk} を用いて表せ。

〔2-2〕　帰無仮説 $H_0 : \pi_X = \pi_Y$ に対する検定統計量を

$$W = \frac{(\hat{\pi}_X - \hat{\pi}_Y)^2}{\hat{V}[\hat{\pi}_X - \hat{\pi}_Y]}, \quad \hat{\pi}_X = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad \hat{\pi}_Y = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

とする。ただし，検定統計量 W の分母は帰無仮説の下での $V[\hat{\pi}_X - \hat{\pi}_Y]$ の未知パラメータを推定値に置き換えた推定量とする。$(X_i, Y_i) = (j, k)$ $(j, k = 0, 1)$ となる人数を n_{jk} としたとき，W を n_{jk} を用いて表せ。

〔3〕　ある疾患の罹患者 200 人それぞれに対し 2 種類の検査 X と Y を行い，その陽性と陰性の結果を表 2 に示す。検査 X と Y の陽性の確率 π_X と π_Y に差があるかどうかを，上問〔2-2〕の検定統計量 W を用いて検定する。

〔3-1〕　帰無仮説の下で，W が近似的に従う分布は何かを述べよ。

〔3-2〕　有意水準 5%で検定した結果を述べよ。

表 2：ある疾患の罹患者 200 人の検査結果

		検査 Y	
		陽性	陰性
検査 X	陽性	134	26
	陰性	16	24

問 2　ある疾患について，試験治療群（T 群）と対照治療群（C 群）のある検査値を比較するランダム化比較試験を考える。各群の被験者の人数はともに n 人であるとし，被験者を表す記号を i として，T 群は $i = 1, \ldots, n$，C 群は $i = n+1, \ldots, 2n$ とする。

被験者 i の治療開始前の検査値を X_i，治療開始 1 年後の検査値を Y_i として，(X_i, Y_i) は次の 2 変量正規分布に従うとする。

$$T\text{群}: \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_T^X \\ \mu_T^Y \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), C\text{群}: \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_C^X \\ \mu_C^Y \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

なお，分散 σ^2 と相関係数 ρ は既知の定数とする。各群の各時点の検査値の標本平均をそれぞれ $\bar{X}_T = \sum_{i=1}^{n} X_i/n$，$\bar{X}_C = \sum_{i=n+1}^{2n} X_i/n$ と，$\bar{Y}_T = \sum_{i=1}^{n} Y_i/n$，$\bar{Y}_C = \sum_{i=n+1}^{2n} Y_i/n$ する。このとき，以下の各問に答えよ。

〔1〕　治療開始 1 年後の検査値 Y_i に注目して，T 群の C 群に対する治療効果を $\delta_d = \mu_T^Y - \mu_C^Y$ とする。δ_d の推定量を $\hat{\delta}_d = \bar{Y}_T - \bar{Y}_C$ とするとき，$\hat{\delta}_d$ の分散を求めよ。

〔2〕　検査値の治療前後の変化量 $Y_i - X_i$ に注目して，T 群の C 群に対する治療効果を $\delta_c = (\mu_T^Y - \mu_T^X) - (\mu_C^Y - \mu_C^X)$ とする。δ_c の推定量を $\hat{\delta}_c = (\bar{Y}_T - \bar{X}_T) - (\bar{Y}_C - \bar{X}_C)$ とするとき，$\hat{\delta}_c$ の分散を求めよ。

〔3〕　Y_i を応答変数，z_i を治療を表す変数，X_i を共変量とする回帰モデル

$$Y_i = \alpha + \delta_a z_i + \rho X_i + \varepsilon_i$$

を考える。処置変数は，T 群では $z_i = 1$，C 群では $z_i = 0$ とし，T 群の C 群に対する治療効果を δ_a とする。α と δ_a は未知のパラメータ，ρ は X_i と Y_i の相関係数（既知の定数），ε_i は互いに独立な誤差項である。

〔3-1〕　最小二乗法による δ_a の推定量が

$$\hat{\delta}_a = c_1 \bar{Y}_T + c_2 \bar{Y}_C + c_3 \bar{X}_T + c_4 \bar{X}_C$$

で与えられるとき，定数 c_1，c_2，c_3，c_4 を求めよ。

〔3-2〕　上問〔3-1〕の推定量 $\hat{\delta}_a$ の分散 $V[\hat{\delta}_a]$ を求めよ。

〔4〕　上問〔1〕,〔2〕,〔3〕の 3 種類の治療効果の推定量について，$E[\hat{\delta}_d] = E[\hat{\delta}_c] = E[\hat{\delta}_a]$ となるための条件を求めよ。

〔5〕　各推定量 $\hat{\delta}_d$，$\hat{\delta}_c$，$\hat{\delta}_a$ の分散について，$\hat{\delta}_d$ の分散 $V[\hat{\delta}_d]$ を基準（分母）とした相対分散 $V[\hat{\delta}_c]/V[\hat{\delta}_d]$，$V[\hat{\delta}_a]/V[\hat{\delta}_d]$ を求め，$V[\hat{\delta}_d]/V[\hat{\delta}_d]$ $(= 1)$，$V[\hat{\delta}_c]/V[\hat{\delta}_d]$，$V[\hat{\delta}_a]/V[\hat{\delta}_d]$ を縦軸，相関係数 ρ（範囲：$0 \leq \rho \leq 1$）を横軸としたグラフを作成せよ。なお，縦軸と横軸には目盛りを入れること。

問 3　試験治療群（T 群）と対照治療群（C 群）の改善割合を比較する臨床試験を考える。以下の各問に答えよ。

〔1〕　T 群と C 群の被験者数をそれぞれ 100 人とする。このうち，T 群では 30 人，C 群では 10 人が，改善の有無が観察されずに欠測データとなった。治療群と欠測の有無に関連があるかどうかを検討するために，治療群と欠測の有無の独立性のカイ二乗検定（有意水準 5%）を用いて，治療群と欠測の有無は独立であるという帰無仮説を検定した結果を述べよ。ただし，カイ二乗統計量は連続補正（イェーツ（Yates）の補正）を施さないものとする。

〔2〕　欠測に対し，多重補完法を適用する。多重補完法によって生成された 3 つのデータセットを表 1 に示す。各補完データセットについて，改善割合の差の分散の推定値（表 1 の (a)，(b)，(c) の値）を求めよ。

表 1：多重補完法によって生成された 3 つの補完データセット

補完データセット	群	改善あり（人）	改善なし（人）	改善割合の差の推定値	改善割合の差の分散の推定値
1	T 群	28	72	0.06	(a)
	C 群	22	78		
2	T 群	30	70	0.07	(b)
	C 群	23	77		
3	T 群	29	71	0.08	(c)
	C 群	21	79		

〔3〕　補完データセット k $(= 1, \ldots, K)$ について，改善割合の差の推定量を $\hat{\theta}_k$，その分散の推定量を $\hat{V}[\hat{\theta}_k]$ とする。そして，K 個の $\hat{\theta}_k$ の平均を $\bar{\theta}$，補完内分散を W，補完間分散を B とする。すなわち，

$$\bar{\theta} = \frac{\sum_{k=1}^{K} \hat{\theta}_k}{K}, \quad W = \frac{\sum_{k=1}^{K} \hat{V}[\hat{\theta}_k]}{K}, \quad B = \frac{\sum_{k=1}^{K} (\hat{\theta}_k - \bar{\theta})^2}{K - 1}$$

である。表 1 の結果から，$\bar{\theta}$ の分散の推定値をルービン（Rubin）のルール

$$\hat{V}[\bar{\theta}] = W + \left(1 + \frac{1}{K}\right) \times B$$

によって求めよ。

〔4〕　T 群の補完データセット k について，改善の有無が観察された人数を N_T，そのうち改善ありが観測された人数を Y_T^o，欠測の人数を M_T，そのうち改善ありと補完された人数を Y_{Tk}^m とする。同様に，C 群のこれらの人数をそれぞれ N_C，Y_C^o，M_C，Y_{Ck}^m とする。補完データセット k から計算される改善割合の差の推定量を

$$\hat{\theta}_k = \frac{Y_T^o + Y_{Tk}^m}{N_T + M_T} - \frac{Y_C^o + Y_{Ck}^m}{N_C + M_C}$$

とする。$Y_T^o = y_T^o$，$Y_C^o = y_C^o$ が与えられた下での，$\hat{\theta}_k$ の条件付き分散を求めよ。ただし，Y_{Tk}^m と Y_{Ck}^m は互いに独立にそれぞれ二項分布 $B(M_T, y_T^o/N_T)$，$B(M_C, y_C^o/N_C)$ に従うものとする。

〔5〕　$Y_T^o = y_T^o$，$Y_C^o = y_C^o$ が与えられた下で，上問〔3〕の補完間分散 B の条件付き期待値を求めよ。

問 4　プラセボ ($j = 0$) に対する，試験治療 40mg ($j = 1$)，80mg ($j = 2$)，120mg ($j = 3$) の 3 用量の効果をランダム化比較試験によって検討した。群 j $(= 0, 1, 2, 3)$ の試験参加者 i $(= 1, \ldots, n_j)$ の結果変数 Y_{ij} は互いに独立に平均 μ_j，分散 σ^2 の正規分布に従うとする。分散 σ^2 は既知とし，各群の結果変数の平均を $\bar{Y}_j = \sum_{i=1}^{n_j} Y_{ij}/n_j$ とする。以下の各問に答えよ。

〔1〕　3 つの帰無仮説

$$H_{01} : \mu_0 = \mu_1, \quad H_{02} : \mu_0 = \mu_2, \quad H_{03} : \mu_0 = \mu_3$$

に対して，それぞれ第 1 種の過誤確率が 2.5%である検定を使用した。各検定の検定統計量は互いに独立であるとしたとき，$\mu_0 = \mu_1 = \mu_2 = \mu_3$ の下で，少なくとも 1 つの帰無仮説が棄却される確率（FWER: familywise error rate）を求めよ。

〔2〕　上問〔1〕の条件の下で，FWER を α 水準以下に制御するために，ボンフェローニ（Bonferroni）手順を使用することにした。この下での各検定の有意水準を求めよ。

〔3〕　一元配置分散分析（有意水準は α）によって，帰無仮説 $H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3$ が棄却されたときに，上問〔1〕の 3 つの帰無仮説 $H_{0j} : \mu_0 = \mu_j$ $(j = 1, 2, 3)$ に対して，それぞれ z 検定（有意水準は α）を使用することにした。各群の人数によらず，FWER が α 水準以下に制御されるための μ_j の条件をすべて答えよ。

〔4〕　上問〔1〕の 3 つの帰無仮説 $H_{0j} : \mu_0 = \mu_j$ $(j = 1, 2, 3)$ に対して，それぞれ検定統計量

$$Z_j = \frac{\bar{Y}_j - \bar{Y}_0}{\sqrt{\sigma^2 \left(\dfrac{1}{n_j} + \dfrac{1}{n_0} \right)}}$$

を用いる。すべての帰無仮説が正しいとき，Z_j と Z_k $(k = 1, 2, 3; j \neq k)$ の相関係数の式を n_0，n_j，n_k を用いて表せ。

〔5〕　帰無仮説 $H_0 : \sum_{j=0}^{3} c_j \mu_j = 0$ に対して，検定統計量

$$Z = \frac{\sum_{j=0}^{3} c_j \bar{Y}_j}{\sqrt{\sigma^2 \sum_{j=0}^{3} (c_j^2 / n_j)}}$$

を使用する。ただし，c_j は解析の目的に応じて設定する係数（対比係数）であり，$\sum_{j=0}^{3} c_j = 0$ を満たすものとする。用量に対して μ_j が直線的に増加するかどうかを検討するためにはどのような対比ベクトル (c_0, c_1, c_2, c_3) を設定すべきか答えよ。ただし，プラセボ群の用量は 0mg とみなし，各群の人数はすべて等しく $n_0 = n_1 = n_2 = n_3$ とする。

〔6〕　上問〔5〕で検討した直線的な用量反応関係に加えて，$(c_0, c_1, c_2, c_3) = (-3, 1, 1, 1)$ に対応する用量反応関係についても検討することにした。帰無仮説が正しいとき，上問〔5〕の対比ベクトルに対応する検定統計量と本問の対比ベクトルに対応する検定統計量の相関係数を求めよ。ただし，各群の人数はすべて等しく $n_0 = n_1 = n_2 = n_3$ とする。

問 5　ある IT 会社では，新卒採用の各応募者に対して 2 段階の試験を実施している。第 1 段階は数学と英語の 2 科目からなる一般教養試験で，第 2 段階は採用に直結した情報の試験である。同社の採用担当の A さんは，第 1 段階の一般教養試験の点数から第 2 段階の情報の試験の点数を予測しようとした。以下では，x_1 を数学の点数，x_2 を英語の点数，y を情報の点数とし，それらの点数はそれぞれ平均 0，分散 1 に標準化されているものとする。また，y と x_1，y と x_2 および x_1 と x_2 間の相関係数をそれぞれ r_{y1}，r_{y2}，r_{12} とする。

A さんが調べたデータの各試験科目間の相関行列は

$$R = \begin{pmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0 \\ 0.6 & 0 & 1 \end{pmatrix} \tag{1}$$

であった。A さんは，英語 (x_2) と情報 (y) 間の相関係数 r_{y2} は 0 であったので，英語は情報の予測に影響を与えないと考え，数学 (x_1) のみを説明変数とする単回帰式を求めたところ $y = 0.6x_1$ となり，このときの決定係数は $R_1^2 = 0.6^2 = 0.36$ であった。しかし同僚の B さんから，x_2 もモデルに入れてみたらとのアドバイスがあり，改めて重回帰式を計算したところ

$$y = 0.8x_1 - 0.4x_2 \tag{2}$$

を得て，決定係数 R_2^2 も大きくなった。A さんは，y との相関が 0 の x_2 をモデルに加えることで x_1 の係数が 0.2 増加して予測力が上がり，モデルの当てはまりがよくなったのが不思議であった。

この分析結果を上司の C さんに報告したところ，C さんから，式 (2) の重回帰式の英語の係数が負であるが，英語力があるほど情報の点数が低いという結果は，IT 業界では英語は不可欠であるので問題であり，A さんの計算が間違っているか，あるいは第 2 段階の情報の試験問題は不適切だったのではないかとの意見があった。

以下の各問に答えよ。なお，一般の相関行列を考察する際は $0 \leq r_{12} < 1$ および $r_{y1} > 0$ とする。

〔1〕　説明変数 x_1，x_2 から y を予測する重回帰式 $y = b_1x_1 + b_2x_2$ の係数 b_1，b_2 の最小二乗法による推定値は，正規方程式

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix}$$

の解として与えられる。正規方程式から b_1 と b_2 を求める計算式を示し，式 (1) の相関行列の数値を代入することで，式 (2) の A さんの計算結果を確認せよ。

〔2〕　A さんの計算では，重回帰式の x_1 の係数 b_1 は単回帰式の x_1 の係数 r_{y1} よりも大きくなった。上問〔1〕の b_1 の計算式を用いて，$b_1 > r_{y1}$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

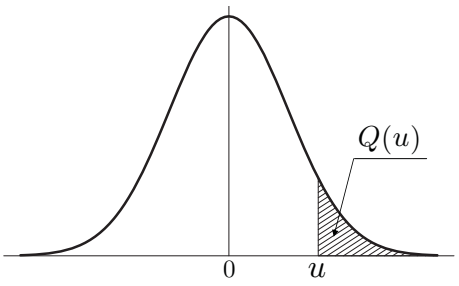
〔3〕　A さんの計算では，$r_{y2} = 0$ であったが重回帰式での x_2 の係数 b_2 の符号が負となった。一般に，$r_{y2} \geq 0$ であるが $b_2 < 0$ となるための r_{12} に関する条件を，r_{y1}，r_{y2} が与えられたとして求めよ。

〔4〕　式 (2) の重回帰式による y の予測値を $\hat{y} = 0.8x_1 - 0.4x_2$ としたとき，$\hat{y}$ の分散および y と $\hat{y}$ の共分散をそれぞれ求めよ。

〔5〕　式 (2) の重回帰式の決定係数 R_2^2 はいくらか。

付　表

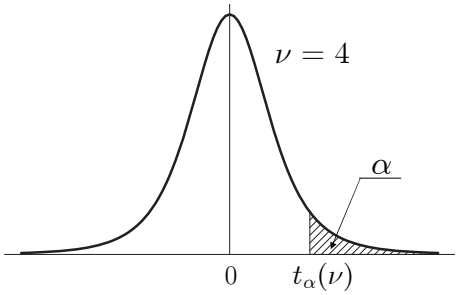
付表 1. 標準正規分布の上側確率



u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

$u = 0.00 \sim 3.99$ に対する，正規分布の上側確率 $Q(u)$ を与える。
例：$u = 1.96$ に対しては，左の見出し 1.9 と上の見出し .06 との交差点で，$Q(u) = .0250$ と読む。表にない u に対しては適宜補間すること。

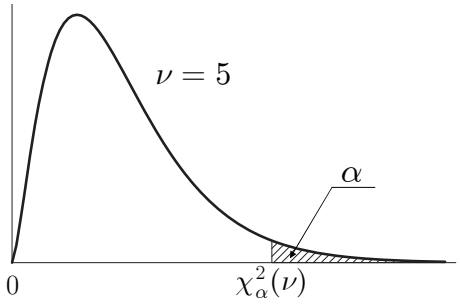
付表 2. t 分布のパーセント点



	α				
ν	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
240	1.285	1.651	1.970	2.342	2.596
∞	1.282	1.645	1.960	2.326	2.576

自由度 ν の t 分布の上側確率 α に対する t の値を $t_\alpha(\nu)$ で表す。
例：自由度 $\nu = 20$ の上側 5%点 ($\alpha = 0.05$) は，$t_{0.05}(20) = 1.725$ である。
表にない自由度に対しては適宜補間すること。

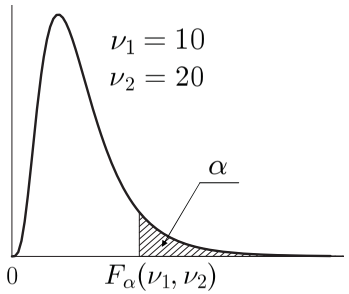
付表 3. カイ二乗分布のパーセント点



	α							
ν	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
35	18.51	20.57	22.47	24.80	46.06	49.80	53.20	57.34
40	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69
50	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
60	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
70	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43
80	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33
90	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12
100	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81
120	86.92	91.57	95.70	100.62	140.23	146.57	152.21	158.95
140	104.03	109.14	113.66	119.03	161.83	168.61	174.65	181.84
160	121.35	126.87	131.76	137.55	183.31	190.52	196.92	204.53
180	138.82	144.74	149.97	156.15	204.70	212.30	219.04	227.06
200	156.43	162.73	168.28	174.84	226.02	233.99	241.06	249.45
240	191.99	198.98	205.14	212.39	268.47	277.14	284.80	293.89

自由度 ν のカイ二乗分布の上側確率 α に対する χ^2 の値を $\chi^2_\alpha(\nu)$ で表す。
例：自由度 $\nu = 20$ の上側 5%点 ($\alpha = 0.05$) は，$\chi^2_{0.05}(20) = 31.41$ である。
表にない自由度に対しては適宜補間すること。

付表 4. F 分布のパーセント点



$\alpha = 0.05$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	15	20	40	60	120	∞
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.619	4.558	4.464	4.431	4.398	4.365
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.845	2.774	2.661	2.621	2.580	2.538
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.403	2.328	2.204	2.160	2.114	2.066
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.203	2.124	1.994	1.946	1.896	1.843
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.089	2.007	1.872	1.822	1.768	1.711
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.015	1.932	1.792	1.740	1.683	1.622
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	1.924	1.839	1.693	1.637	1.577	1.509
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.836	1.748	1.594	1.534	1.467	1.389
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.750	1.659	1.495	1.429	1.352	1.254

$\alpha = 0.025$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	15	20	40	60	120	∞
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.428	6.329	6.175	6.123	6.069	6.015
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.522	3.419	3.255	3.198	3.140	3.080
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.862	2.756	2.585	2.524	2.461	2.395
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.573	2.464	2.287	2.223	2.156	2.085
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.411	2.300	2.118	2.052	1.981	1.906
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.307	2.195	2.009	1.940	1.866	1.787
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.182	2.068	1.875	1.803	1.724	1.637
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.061	1.944	1.744	1.667	1.581	1.482
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	1.945	1.825	1.614	1.530	1.433	1.310

自由度 (ν_1, ν_2) の F 分布の上側確率 α に対する F の値を $F_\alpha(\nu_1, \nu_2)$ で表す。
例：自由度 $\nu_1 = 5$, $\nu_2 = 20$ の上側 5% 点 $(\alpha = 0.05)$ は，$F_{0.05}(5, 20) = 2.711$ である。
表にない自由度に対しては適宜補間すること。

付表 5. 指数関数と常用対数

指数関数			
x	e^x	x	e^x
0.01	1.0101	0.51	1.6653
0.02	1.0202	0.52	1.6820
0.03	1.0305	0.53	1.6989
0.04	1.0408	0.54	1.7160
0.05	1.0513	0.55	1.7333
0.06	1.0618	0.56	1.7507
0.07	1.0725	0.57	1.7683
0.08	1.0833	0.58	1.7860
0.09	1.0942	0.59	1.8040
0.10	1.1052	0.60	1.8221
0.11	1.1163	0.61	1.8404
0.12	1.1275	0.62	1.8589
0.13	1.1388	0.63	1.8776
0.14	1.1503	0.64	1.8965
0.15	1.1618	0.65	1.9155
0.16	1.1735	0.66	1.9348
0.17	1.1853	0.67	1.9542
0.18	1.1972	0.68	1.9739
0.19	1.2092	0.69	1.9937
0.20	1.2214	0.70	2.0138
0.21	1.2337	0.71	2.0340
0.22	1.2461	0.72	2.0544
0.23	1.2586	0.73	2.0751
0.24	1.2712	0.74	2.0959
0.25	1.2840	0.75	2.1170
0.26	1.2969	0.76	2.1383
0.27	1.3100	0.77	2.1598
0.28	1.3231	0.78	2.1815
0.29	1.3364	0.79	2.2034
0.30	1.3499	0.80	2.2255
0.31	1.3634	0.81	2.2479
0.32	1.3771	0.82	2.2705
0.33	1.3910	0.83	2.2933
0.34	1.4049	0.84	2.3164
0.35	1.4191	0.85	2.3396
0.36	1.4333	0.86	2.3632
0.37	1.4477	0.87	2.3869
0.38	1.4623	0.88	2.4109
0.39	1.4770	0.89	2.4351
0.40	1.4918	0.90	2.4596
0.41	1.5068	0.91	2.4843
0.42	1.5220	0.92	2.5093
0.43	1.5373	0.93	2.5345
0.44	1.5527	0.94	2.5600
0.45	1.5683	0.95	2.5857
0.46	1.5841	0.96	2.6117
0.47	1.6000	0.97	2.6379
0.48	1.6161	0.98	2.6645
0.49	1.6323	0.99	2.6912
0.50	1.6487	1.00	2.7183

常用対数			
x	$\log_{10} x$	x	$\log_{10} x$
0.1	-1.0000	5.1	0.7076
0.2	-0.6990	5.2	0.7160
0.3	-0.5229	5.3	0.7243
0.4	-0.3979	5.4	0.7324
0.5	-0.3010	5.5	0.7404
0.6	-0.2218	5.6	0.7482
0.7	-0.1549	5.7	0.7559
0.8	-0.0969	5.8	0.7634
0.9	-0.0458	5.9	0.7709
1.0	0.0000	6.0	0.7782
1.1	0.0414	6.1	0.7853
1.2	0.0792	6.2	0.7924
1.3	0.1139	6.3	0.7993
1.4	0.1461	6.4	0.8062
1.5	0.1761	6.5	0.8129
1.6	0.2041	6.6	0.8195
1.7	0.2304	6.7	0.8261
1.8	0.2553	6.8	0.8325
1.9	0.2788	6.9	0.8388
2.0	0.3010	7.0	0.8451
2.1	0.3222	7.1	0.8513
2.2	0.3424	7.2	0.8573
2.3	0.3617	7.3	0.8633
2.4	0.3802	7.4	0.8692
2.5	0.3979	7.5	0.8751
2.6	0.4150	7.6	0.8808
2.7	0.4314	7.7	0.8865
2.8	0.4472	7.8	0.8921
2.9	0.4624	7.9	0.8976
3.0	0.4771	8.0	0.9031
3.1	0.4914	8.1	0.9085
3.2	0.5051	8.2	0.9138
3.3	0.5185	8.3	0.9191
3.4	0.5315	8.4	0.9243
3.5	0.5441	8.5	0.9294
3.6	0.5563	8.6	0.9345
3.7	0.5682	8.7	0.9395
3.8	0.5798	8.8	0.9445
3.9	0.5911	8.9	0.9494
4.0	0.6021	9.0	0.9542
4.1	0.6128	9.1	0.9590
4.2	0.6232	9.2	0.9638
4.3	0.6335	9.3	0.9685
4.4	0.6435	9.4	0.9731
4.5	0.6532	9.5	0.9777
4.6	0.6628	9.6	0.9823
4.7	0.6721	9.7	0.9868
4.8	0.6812	9.8	0.9912
4.9	0.6902	9.9	0.9956
5.0	0.6990	10.0	1.0000

注: 常用対数を自然対数に直すには 2.3026 をかければよい。

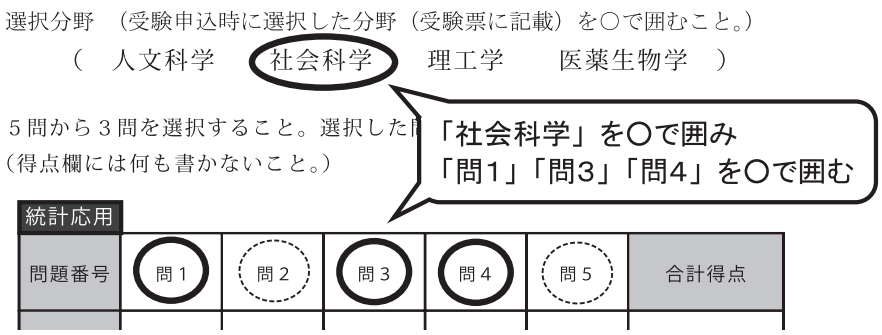
【解答冊子記入例】

- 注意事項 6 ⑥：〔解答冊子各ページ先頭の記入例〕

（例）問 1 を解答する場合

- 注意事項 6 ⑦：〔解答冊子表紙選択分野・選択問題の記入例〕

（例）**社会科学** 分野の**問 1**，**問 3**，**問 4** を選択し，解答する場合



一般財団法人　統計質保証推進協会
統計検定センター

〒101-0051 東京都千代田区神田神保町３丁目６番
URL **http://www.toukei-kentei.jp**

2024.11