統計検定
Japan Statistical Society Certificate

1級
統計応用

2019 年 11 月 24 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**，**15〜25 ページ**は**社会科学**，**27〜37 ページ**は**理工学**，**39〜49 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**しなさい。

⑤ **問**ごと（**問 1**，**問 2**，...）に解答のページを改めなさい。

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 51 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

問 1　S 大学の入試では，受験者の得点により合格者を決定している。ある年，受験者全体の得点の上位半分を合格とした。また，受験者全体での得点の上位 10%の合格者は入学しなかった。この年，A 君は S 大学を受験し合格した。A 君の得点は受験者全体の中で偏差値が 54 であった。

受験者全体の得点は正規分布 $N(100, 20^2)$ と見なし，以下の各問に答えよ。なお，標準正規分布の確率密度関数は $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left[-\dfrac{z^2}{2}\right]$ である。

〔1〕　A 君の得点はいくらで，それは受験者全体の中で上位何%に位置する得点であるかを示せ。また，A 君の得点は合格者および入学者の中で上位何%に位置する得点であるかをそれぞれ示せ。

〔2〕　入学者の最低点と最高点はいくらか。

〔3〕　合格者の得点の分布の確率密度関数を，標準正規分布の確率密度関数 $\varphi(z)$ を用いて示せ。

〔4〕　合格者の得点の平均および分散はそれぞれいくらか。

問２　ある不動産会社が，ある駅近辺のマンションの分類のため，A〜F の６棟のマンションの築年数（単位：年）と駅からの所要時間（単位：分）を調査したところ，表１のような結果を得た。また，表２は表１から求めたユークリッド距離によるマンション間の距離行列である。これらの表を基にしたクラスター分析について，以下の各問に答えよ。

表 1: 築年数と駅からの所要時間

マンション	築年数	所要時間
A	1	3
B	2	2
C	3	4
D	7	9
E	8	7
F	9	9

表 2: 表１から作成したユークリッド距離による各マンション間の距離行列

	A	B	C	D	E	F
A	0.00	1.41	2.24	8.49	8.06	10.00
B	1.41	0.00	2.24	8.60	7.81	9.90
C	2.24	2.24	0.00	6.40	5.83	7.81
D	8.49	8.60	6.40	0.00	2.24	2.00
E	8.06	7.81	5.83	2.24	0.00	2.24
F	10.00	9.90	7.81	2.00	2.24	0.00

表１のデータに対し，階層的クラスター分析を適用することでクラスターを求める。

〔1〕　最短距離法（single-linkage method）を用いた階層的クラスター分析を行い，デンドログラムを作成せよ。

〔2〕　２つのクラスターに分けたいとき，デンドログラムをどの距離の区間で分ければよいか述べよ。

次に，表 1 のデータに対して非階層的クラスター分析を行う。ここでは，初期値としてデータセットの中からランダムに k 点の代表点を選んでクラスターを決定していく k-means 法を適用する。

〔3〕 初期割り振りとしてマンション D をクラスター1，マンション E をクラスター2 の代表点に割り振った状態を考える。クラスター代表点を 1 回更新したときの，各クラスターの代表点座標を計算し，更新後の各クラスターに属するマンションの記号をそれぞれ答えよ。

〔4〕 k-means 法によるクラスター分析では初期値依存性があることが知られている。初期値依存性とは何かについて簡潔に説明せよ。また，初期値依存性によって誤った結論になるのを避けるための対策について述べよ。

問 3　項目反応理論における 2 パラメータロジスティックモデルでは，パラメータ θ の参加者が項目 j に正答する確率を表す項目反応関数 (item response function) が

$$P_j(\theta) = \frac{1}{1 + \exp[-a_j(\theta - b_j)]}$$

によって与えられる（$\exp[x] = e^x$ である）。ここで a_j と b_j は項目パラメータである。データとして与えられるのは，各項目について正答のとき 1, 誤答のとき 0 の値をとる 2 値観測変数とする。このとき，以下の各問に答えよ。

〔1〕　項目パラメータが次のように与えられる 4 つの項目がある。

	a_j	b_j
項目 1	1.0	2.0
項目 2	1.0	1.0
項目 3	0.5	2.0
項目 4	0.5	−2.0

$\theta = 0$ の参加者にとって正答確率が等しい 2 つの項目はどれとどれかを示せ。

〔2〕　参加者パラメータ θ が参加者集団において標準正規分布に従うとき，この参加者集団にとって，上問〔1〕の 4 つの項目の中で最も正答しやすいと考えられる項目はどれかを求めよ。

〔3〕　項目反応関数の θ に関する偏導関数は，$P_j(\theta) = 0.5$ である点においてどのようになるかを項目パラメータの式で示せ。

〔4〕　項目 j について

$$f_j(\theta) = \frac{\left(\dfrac{\partial P_j(\theta)}{\partial \theta}\right)^2}{P_j(\theta)\{1 - P_j(\theta)\}}$$

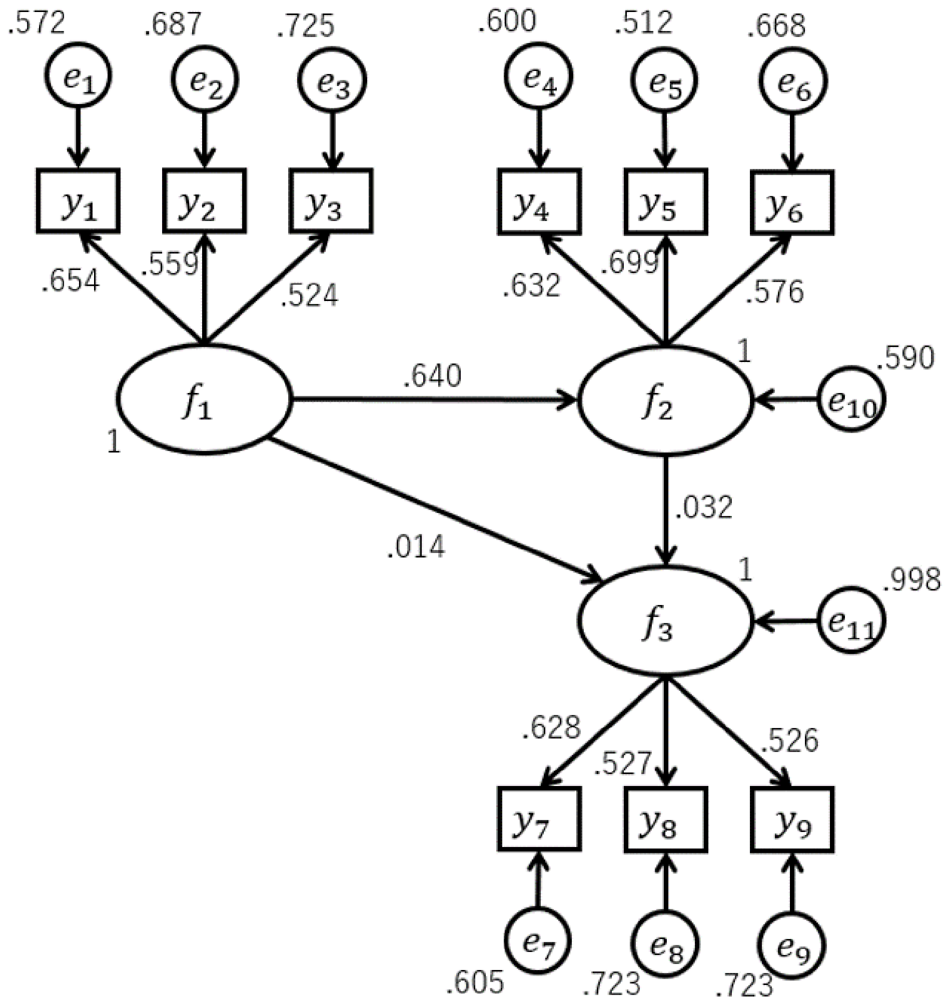
によって与えられる関数の名称を述べ，これが何を表すかを説明せよ。

〔5〕　パラメータ c_j を追加した 3 パラメータロジスティックモデル

$$P_j(\theta) = c_j + \frac{1 - c_j}{1 + \exp[-a_j(\theta - b_j)]}$$

におけるパラメータ c_j の意味を述べよ。

問 4　次の図は，あるソフトウェアを用いて，あるデータについて構造方程式モデリング（共分散構造分析）を行った結果から得られたパス図と各パラメータの推定値である。図中に示された推定値に基づき，以下の各問に答えよ。



〔1〕　この結果は標準解と非標準解のどちらと考えられるか。理由とともに述べよ。

〔2〕　上問〔1〕で答えた解における，f_1 から f_2 への直接効果，間接効果，総合効果の推定値をそれぞれ求めよ。

〔3〕　上問〔1〕で答えた解における，f_1 から f_3 への直接効果，間接効果，総合効果の推定値をそれぞれ求めよ。

〔4〕　このモデルにおける潜在変数の構造方程式は，$\boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$ とするとき，

$$\boldsymbol{f} = A\boldsymbol{f} + \boldsymbol{e}$$

と表現できる。行列 A とベクトル $\boldsymbol{e}$ の要素を具体的に示せ。ただし，必要な場合には $\boldsymbol{f}$ および $\boldsymbol{e}$ の要素は記号のままでよい。

〔5〕　構造方程式モデリングでは，同値モデルの存在が問題となることがある。同値モデルとは何かを説明せよ。

問 5　ABO 血液型の分布は O 型，A 型，B 型，AB 型の比率で示され，この比率は国や地域によって違いが見られる。日本のある地域 C から無作為に抽出した 100 人を調べたところ，血液型の分布は表 1 のようになった。以下の各問に答えよ。

表 1: 地域 C の血液型分布（観測度数）

血液型	O 型	A 型	B 型	AB 型	合計
観測度数	24	48	16	12	100

〔1〕　日本人の ABO 血液型は

$$\text{O 型} : \text{A 型} : \text{B 型} : \text{AB 型} = 3 : 4 : 2 : 1 \tag{1}$$

の比率で分布すると言われている。帰無仮説を 式 (1) の比率とし，表 1 について適合度のカイ 2 乗検定を有意水準 5%で行い，その結果を述べよ。

〔2〕　血液型の分布の観測度数が，k を自然数として $6k$，$12k$，$4k$，$3k$ であったとしたとき（表 1 では $k = 4$），適合度のカイ 2 乗検定が有意水準 5%で有意になる最小の k はいくらか。

〔3〕　一般に，適合度のカイ 2 乗検定統計量は，近似的にカイ 2 乗分布に従うことからその名があるが，その統計量が近似的にカイ 2 乗分布に従う根拠は何かを詳細に述べよ。厳密に証明する必要はない。

問題は次ページに続く。

〔4〕　ABO 血液型は，親から受け継いだ 3 つの遺伝子 O，A，B の組合せによって決まることが知られていて，表 2 のように血液型が決まる。これより，遺伝子 O，A，B はそれぞれ r，p，q $(r+p+q=1)$ の比率で分布しているとすると，各血液型の比率は表 2 の最後の行に示したようになる。

表 2: 遺伝子を考慮した血液型分布

血液型	O 型	A 型	B 型	AB 型
遺伝子型	OO	AA AO OA	BB BO OB	AB BA
比率	r^2	p^2+2pr	q^2+2qr	$2pq$

全観測度数を N とし，各血液型の観測度数をそれぞれ n_O，n_A，n_B，n_{AB}，各遺伝子型の度数を f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} とする（f_{AO}，f_{BO}，f_{AB} はそれぞれ AO と OA，BO と OB，AB と BA の合計度数である）。このとき，$f_{OO} = n_O$，$f_{AA}+f_{AO}=n_A$，$f_{BB}+f_{BO}=n_B$，$f_{AB}=n_{AB}$ であり，f_{AA}，f_{AO}，f_{BB}，f_{BO} は実際は観測されない度数である。

比率 r，p，q の最尤推定値を求める。度数 f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} に基づく尤度関数は，r，p，q に依存しない定数を無視すると

$$L(r,\,p,\,q) \propto (r^2)^{f_{OO}}(p^2)^{f_{AA}}(2pr)^{f_{AO}}(q^2)^{f_{BB}}(2qr)^{f_{BO}}(2pq)^{f_{AB}}$$

となる。次の (i) および (ii) に答え，最尤推定値を求める数値計算の反復法を構築せよ。ただし実際に数値を求める必要はない。

(i)　ラグランジュの未定乗数を λ とした

$$Q = \log L(r,\,p,\,q) - \lambda(r+p+q-1)$$

を r，p，q でそれぞれ偏微分して 0 と置き，$L(r,\,p,\,q)$ を最大化する r，p，q の値を求める式を示せ。

(ii)　上記 (i) で求めた r，p，q を用いて度数 f_{AA}，f_{AO}，f_{BO}，f_{BB} の期待値を求める式を示せ。

社会科学

問 1　大きさ N の有限母集団が L 個の部分母集団（層）に分割され，各層の大きさを N_h $(h = 1, \ldots, L)$ とする $(N = \sum_{h=1}^{L} N_h$ である)。ある測定項目 Y について，第 h 層における第 i 番目の個体の値を y_{hi} $(i = 1, \ldots, N_h; h = 1, \ldots, L)$ と書き，各層の母平均と母分散を

$$\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}, \quad S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$$

で定める。第 h 層の大きさの相対比率を $W_h = \dfrac{N_h}{N}$ とし，全体の母集団平均 $\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h$ を推定する目的で，層別に標本を抽出することを考える。

各層の標本の大きさを n_h として，全体で大きさ $n = \sum_{h=1}^{L} n_h$ の標本を抽出する。各層では，他の層とは独立に非復元単純無作為抽出を行い，第 h 層での標本平均を $\bar{y}_h = \dfrac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ とする。また，$w_h = \dfrac{n_h}{n}$, $f_h = \dfrac{n_h}{N_h}$ と記号を定める。このとき，以下の各問に答えよ。ただし，$\bar{y}_h$ の期待値と分散がそれぞれ

$$E[\bar{y}_h] = \bar{Y}_h, \quad V[\bar{y}_h] = \frac{1 - f_h}{n_h} S_h^2 \quad (h = 1, \ldots, L)$$

となることを用いてもよい。

〔1〕　層化推定量 $\hat{Y} = \sum_{h=1}^{L} W_h \bar{y}_h$ の期待値 $E[\hat{Y}]$ と分散 $V[\hat{Y}]$ を求めよ。

〔2〕　各層の標本を大きさ $n_h = W_h n$ で割り当てたときの層化推定量を $\hat{Y}_{(1)}$ と表す。分散 $V[\hat{Y}_{(1)}]$ を求めよ。

〔3〕　標本の大きさ n が与えられたとき，層化推定量の分散を最小にするような標本の配分 n_h を求めよ。

〔4〕　上問〔3〕の配分 n_h を用いた層化推定量を $\hat{Y}_{(2)}$ と表す。分散 $V[\hat{Y}_{(2)}]$ を求めよ。

〔5〕　層の情報を用いない大きさ n の非復元無作為標本 $y_1, \ldots, y_n$ から求めた推定量を $\hat{Y}_{(0)} = \bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ と書く。有限母集団の修正項が $1 - f_h \approx 1$ と無視できるほど各層の大きさ N_h が大きい場合，関係式

$$V[\hat{Y}_{(2)}] \leq V[\hat{Y}_{(1)}] \leq V[\hat{Y}_{(0)}]$$

が成り立つことを示せ。

問 2　ある大きな母集団における世帯所得の分布を考える。この分布に従う確率変数 $X \geq 0$ は連続型で，その累積分布関数を $F(x)$，確率密度関数を $f(x)$ とし，X の期待値を $\mu = E[X] < \infty$ とする。ここでは，$u = F(x)$ が $0 < F(x) < 1$ となる x で狭義単調増加，すなわち $0 < u < 1$ で逆関数 $x = F^{-1}(u)$ が存在する場合のみを考える。

この分布のローレンツ曲線 $L(u)$ は，$0 \leq u \leq 1$ で

$$L(u) = \frac{\displaystyle\int_0^u F^{-1}(t)dt}{\displaystyle\int_0^1 F^{-1}(t)dt} = \frac{\displaystyle\int_0^{F^{-1}(u)} yf(y)dy}{\mu} \tag{1}$$

により定義される（$u = 1$ のときは，式 (1) の最右辺の分子における積分範囲の $F^{-1}(u)$ は ∞ もしくは $\displaystyle\lim_{u \uparrow 1} F^{-1}(u)$ と考える）。また，ジニ係数 G は 45 度線とローレンツ曲線で囲まれた部分の面積の 2 倍として定義される。すなわち，

$$G = 2\int_0^1 \{u - L(u)\}du$$

である。このとき，以下の各問に答えよ。なお以下では，上で述べた条件を満たす分布のみを考える。

〔1〕　ローレンツ曲線およびジニ係数は所得分布のどのような特徴を表すかを簡潔に述べよ。

〔2〕　確率変数 $U = F(X)$ は区間 $(0, 1)$ 上の一様分布に従うことを示せ。

〔3〕　$\sigma > 0, \alpha > 1$ に対して

$$F(x) = \begin{cases} 1 - \left(\dfrac{x}{\sigma}\right)^{-\alpha} & (x > \sigma) \\ 0 & (0 \leq x \leq \sigma) \end{cases}$$

を累積分布関数として持つ分布をパレート分布といい，$\mathrm{Pareto}(\sigma, \alpha)$ と書く。パレート分布 $\mathrm{Pareto}(\sigma, \alpha)$ のローレンツ曲線 $L(u)\ (0 \leq u \leq 1)$ を求めよ。またそれを利用してジニ係数は $G = \dfrac{1}{2\alpha - 1}$ となることを示せ。

〔4〕 一般に，2 つの分布 F_1, F_2 の累積分布関数を $F_1(x), F_2(x)$ とし，ローレンツ曲線をそれぞれ $L_1(u), L_2(u)$ としたとき，すべての $0 \leq u \leq 1$ に対して $L_1(u) \geq L_2(u)$ であり，かつある $0 < u < 1$ において $L_1(u) > L_2(u)$ となるとき，F_1 は F_2 をローレンツ優越するという。

上問〔2〕で定義したパレート分布に対し，$\mathrm{Pareto}(\sigma_1, \alpha_1)$ が $\mathrm{Pareto}(\sigma_2, \alpha_2)$ をローレンツ優越するための必要十分条件を $\sigma_1, \alpha_1, \sigma_2, \alpha_2$ を用いて表せ。

〔5〕 ある分布の期待値が $\mu > 0$ であり，ローレンツ曲線が

$$L(u) = u + (1-u)\log(1-u) \quad (0 \leq u \leq 1)$$

であるとする。このとき，この分布の累積分布関数 $F(x)$ を求めよ。なお，ここで $\log$ は自然対数である。

問 3　時系列データ $(\ldots, X_{-1}, X_0, X_1, \ldots, X_n, \ldots)$ は 1 次の自己回帰（AR(1)）モデル

$$X_t = \phi X_{t-1} + \epsilon_t \quad (t = \ldots, -1, 0, 1, \ldots) \tag{1}$$

に従うとする。ここで ϵ_t は互いに独立に $N(0, \sigma^2)$ に従う確率変動項であり，定常性の条件 $|\phi| < 1$ を仮定する。$(X_1, \ldots, X_n)$ につき以下の各問に答えよ。

〔1〕　モデル (1) における $(X_1, \ldots, X_n)$ の自己共分散行列 $T = \{\tau_{ij}\}$ の各成分は

$$\tau_{ij} = \frac{\sigma^2}{1-\phi^2}\phi^{|i-j|} \quad (i, j = 1, \ldots, n)$$

で与えられ，自己相関行列は $R = \{\rho_{ij}\} = \{\phi^{|i-j|}\}$ となることを示せ。

〔2〕　n 次対称行列 $A = \{a_{ij}\}$ を

$$a_{ij} = \begin{cases} 1 & (i = j = 1, i = j = n) \\ 1+\phi^2 & (i = j = 2, \ldots, n-1) \\ -\phi & (|i-j| = 1) \\ 0 & (|i-j| \geq 2) \end{cases}$$

とする。たとえば $n = 4$ では

$$A = \begin{pmatrix} 1 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{pmatrix}$$

である。

一般の n および上問〔1〕の行列 T に対し，$\dfrac{1}{\sigma^2}T$ の逆行列は A で与えられることを示せ。また，A の行列式 $|A|$ および R の行列式 $|R|$ の値を求めよ。

〔3〕　$\boldsymbol{x} = (x_1, \ldots, x_n)'$ を n 次ベクトルとしたとき（プライム $(')$ は転置を表す），上問〔2〕の行列 A に関する 2 次形式 $Q_A = \boldsymbol{x}'A\boldsymbol{x}$ を $x_1, \ldots, x_n$ を用いて書き下し，$|\phi| < 1$ のとき，Q_A はすべての $\boldsymbol{x} \neq \boldsymbol{0}$（成分がすべて 0 のベクトル）に対して常に正であること，すなわち A は正定値であることを示せ。

〔4〕　上問〔2〕の行列 A の $(1,1)$ 要素の 1 のみを ϕ^2 に変えた行列を B とする。たとえば $n=4$ の場合は

$$B = \begin{pmatrix} \phi^2 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{pmatrix}$$

である。一般の n について，B に関する 2 次形式 $Q_B = \boldsymbol{x}'B\boldsymbol{x}$ はすべての $\boldsymbol{x}$ に対して ϕ の値によらず非負となること，すなわち Q_B は非負定値であることを示せ。また，$Q_B = 0$ となる $\boldsymbol{x}$ (ただし，$\boldsymbol{x} \neq \mathbf{0}$) はどのようなベクトルであるか。

〔5〕　モデル (1) における自己回帰係数 ϕ が既知もしくはきわめて精度良く推定されているが誤差分散 σ^2 は未知であるとき，σ^2 の 95 %信頼区間の構成法を示せ。

問 4　あるコーヒーショップチェーンでは，各店舗から報告されるお客さんからのクレーム情報を収集して分析し，よりよい接客につなげようとしている。表 1 はある日に全国の店舗から無作為に抽出した 100 店舗のクレームの件数の度数とそれらの平均および分散である。この表に関する以下の令央君と和美さんの会話に関する各問に答えよ。なお，パラメータ λ のポアソン分布に従う確率変数 X の確率関数は

$$f(x;\lambda) = P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda} \quad (x = 0, 1, 2, \dots) \tag{1}$$

である。

表 1: クレーム数の度数分布と平均，分散

クレーム数	0	1	2	3	4	5	6	7	計	平均	分散
度数	22	23	26	18	6	4	1	0	100	1.79	2.03

令央：クレームは稀な事象だからポアソン分布が当てはまると思うよ。ポアソン分布のパラメータ λ の最尤推定値は標本平均だから，データから値を求めると $\hat{\lambda} = \bar{x} = 1.79$ だ。これをパラメータ値とするポアソン分布を折れ線に表示して，表 1 の度数の棒グラフに当てはめると図 1 になったよ。

和美：なんだか当てはまりが悪いわね。クレーム数の分布はポアソン分布とはちょっと違うんじゃないかしら。ポアソン分布だと平均と分散が等しいはずだけど，分散は 2.03 と平均の 1.79 よりも大きいし。噂だけど本当はクレームがあったのにそれがなかったと報告している店舗があるみたいよ。

令央：では，クレームがあったのにそれを 0 と報告した店舗の割合を ω として，ゼロ度数が多いゼロ過剰なポアソン分布を当てはめてみよう。計算の結果今度は λ の推定値が $\hat{\lambda} = 1.98$ になったので分布形をグラフにしてみると図 2 のようになったよ。

和美：当てはまりが格段に良くなったわ。クレームがあったのにそれを 0 としたショップの割合 ω はどのくらいなのかしら。そういう店舗には正直に報告してもらうようお願いしなくてはいけないわね。クレーム情報は接客の向上のために有用ですもの。

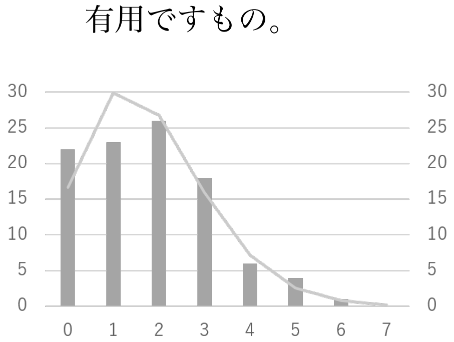


図 1: ポアソン分布の当てはめ

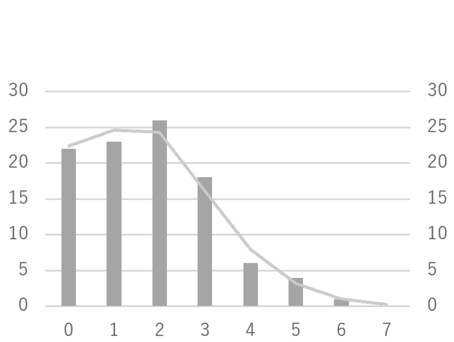


図 2: ゼロ過剰なポアソン分布の当てはめ

〔1〕　パラメータ λ のポアソン分布の期待値と分散はともに λ であることを示せ。

〔2〕　パラメータ λ のポアソン分布からの大きさ n の無作為標本が $x_1, \ldots, x_n$ と得られたときのパラメータ λ の最尤推定値は標本平均 $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ であることを示せ。

〔3〕　$x = k$ となる度数を f_k とし，表 1 のような観測度数分布 $(f_0, f_1, \ldots, f_K)$ に対するポアソン分布の適合度を統計的に評価する適合度のカイ 2 乗検定について述べよ。ただし，実際に計算する必要はない。

〔4〕　ゼロ過剰パラメータを ω としたときのゼロ過剰なポアソン（Zero-Inflated Poisson = ZIP）分布の確率関数は

$$g(x; \lambda, \omega) = \begin{cases} \omega + (1-\omega) f(x; \lambda) & (x = 0) \\ (1-\omega) f(x; \lambda) & (x \geq 1) \end{cases} \tag{2}$$

で与えられる。ここで $f(x; \lambda)$ は式 (1) で与えられるポアソン分布の確率関数である。ZIP 分布からの大きさ n の無作為標本 $x_1, \ldots, x_n$ が与えられたとき，以下の手順に基づき，繰り返し計算による λ の最尤推定法を示せ。

(i)　$x = 0$ の観測度数を f_0 とし（表 1 では $f_0 = 22$），$A = x_1 + \cdots + x_n$ と置く。また，便宜的にゼロ過剰部分の（観測されない）度数を m とする。すなわちポアソン分布部分の度数は $n - m$ である。このとき，n, m, A を用いて λ および ω の対数尤度関数 $l(\lambda, \omega)$ 求めよ。

(ii)　$l(\lambda, \omega)$ を λ および ω で偏微分して 0 と置くことにより

$$\begin{cases} \lambda = h_1(n, m, A) \\ \omega = h_2(n, m) \end{cases}$$

なる関係式を導け。

(iii)　m と λ の関係式

$$m = h_3(n, f_0, \lambda)$$

を導き，$\lambda^{(0)}$ を初期値として，繰り返し計算式

$$m^{(t)} = h_3(n, f_0, \lambda^{(t-1)})$$
$$\lambda^{(t)} = h_1(n, m^{(t)}, A)$$

を求めよ。

〔5〕　表 1 のデータに基づく，式 (2) の ZIP 分布におけるゼロ過剰パラメータ ω の推定値はいくらか。また，観測されたクレーム数が 0 の度数 22 のうちで，クレームがあったにもかかわらずそれを 0 と報告したショップの割合はいくらか。

問 5　ABO 血液型の分布は O 型，A 型，B 型，AB 型の比率で示され，この比率は国や地域によって違いが見られる。日本のある地域 C から無作為に抽出した 100 人を調べたところ，血液型の分布は表 1 のようになった。以下の各問に答えよ。

表 1: 地域 C の血液型分布（観測度数）

血液型	O 型	A 型	B 型	AB 型	合計
観測度数	24	48	16	12	100

〔1〕　日本人の ABO 血液型は

$$\text{O 型} : \text{A 型} : \text{B 型} : \text{AB 型} = 3 : 4 : 2 : 1 \tag{1}$$

の比率で分布すると言われている。帰無仮説を 式 (1) の比率とし，表 1 について適合度のカイ 2 乗検定を有意水準 5%で行い，その結果を述べよ。

〔2〕　血液型の分布の観測度数が，k を自然数として $6k$，$12k$，$4k$，$3k$ であったとしたとき（表 1 では $k = 4$），適合度のカイ 2 乗検定が有意水準 5%で有意になる最小の k はいくらか。

〔3〕　一般に，適合度のカイ 2 乗検定統計量は，近似的にカイ 2 乗分布に従うことからその名があるが，その統計量が近似的にカイ 2 乗分布に従う根拠は何かを詳細に述べよ。厳密に証明する必要はない。

問題は次ページに続く。

〔4〕 ABO 血液型は，親から受け継いだ 3 つの遺伝子 O，A，B の組合せによって決まることが知られていて，表 2 のように血液型が決まる。これより，遺伝子 O，A，B はそれぞれ r，p，q $(r+p+q=1)$ の比率で分布しているとすると，各血液型の比率は表 2 の最後の行に示したようになる。

表 2: 遺伝子を考慮した血液型分布

血液型	O 型	A 型	B 型	AB 型
遺伝子型	OO	AA AO OA	BB BO OB	AB BA
比率	r^2	p^2+2pr	q^2+2qr	$2pq$

全観測度数を N とし，各血液型の観測度数をそれぞれ n_O，n_A，n_B，n_{AB}，各遺伝子型の度数を f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} とする（f_{AO}，f_{BO}，f_{AB} はそれぞれ AO と OA，BO と OB，AB と BA の合計度数である）。このとき，$f_{OO} = n_O$，$f_{AA}+f_{AO}=n_A$，$f_{BB}+f_{BO}=n_B$，$f_{AB}=n_{AB}$ であり，f_{AA}，f_{AO}，f_{BB}，f_{BO} は実際は観測されない度数である。

比率 r，p，q の最尤推定値を求める。度数 f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} に基づく尤度関数は，r，p，q に依存しない定数を無視すると

$$L(r,\,p,\,q) \propto (r^2)^{f_{OO}}(p^2)^{f_{AA}}(2pr)^{f_{AO}}(q^2)^{f_{BB}}(2qr)^{f_{BO}}(2pq)^{f_{AB}}$$

となる。次の (i) および (ii) に答え，最尤推定値を求める数値計算の反復法を構築せよ。ただし実際に数値を求める必要はない。

(i) ラグランジュの未定乗数を λ とした

$$Q = \log L(r,\,p,\,q) - \lambda(r+p+q-1)$$

を r，p，q でそれぞれ偏微分して 0 と置き，$L(r,\,p,\,q)$ を最大化する r，p，q の値を求める式を示せ。

(ii) 上記 (i) で求めた r，p，q を用いて度数 f_{AA}，f_{AO}，f_{BO}，f_{BB} の期待値を求める式を示せ。

理工学

問 1　ある工業製品の寿命を表す連続型の確率変数を T とし $(T \geq 0)$，その累積分布関数を $F(t) = P(T \leq t)$，生存関数を $S(t) = P(T > t) = 1 - F(t)$，確率密度関数を $f(t) = \frac{d}{dt}F(t)$ とする。また，ハザード関数を $h(t) = \frac{f(t)}{S(t)}$ とし，累積ハザード関数を $H(t) = \int_0^t h(s)ds$ とする。以下では $F(t)$ および $H(t)$ は適当な回数微分可能であり，$E[T] < \infty$ とする。このとき以下の各問に答えよ。

〔1〕　T の期待値を $E[T] = \int_0^\infty tf(t)dt$ とするとき，

$$E[T] = \int_0^\infty S(t)dt$$

となることを示せ。

〔2〕　この製品が時点 t で稼働しているときの余命 $T - t$ の期待値（平均余命関数）を

$$m(t) = E[T - t | T > t]$$

とする。このとき

$$m(t) = \frac{\int_t^\infty S(x)dx}{S(t)}$$

および

$$m(t) = \int_0^\infty \exp[H(t) - H(t + x)]dx$$

であることを示せ。また，$m'(x)$ を $m(x)$ の導関数とするとき

$$S(t) = \exp\left[-\int_0^t \frac{1 + m'(x)}{m(x)}dx\right]$$

であることを示せ。

〔3〕　ハザード関数 $h(t)$ が t の増加（非減少）関数であるとき寿命分布は IFR（increasing failure rate）であるといい，それが t の減少（非増加）関数のとき DFR（decreasing failure rate）であるという。

(1)　ハザード関数 $h(t)$ は寿命のどのような性質を意味するかを述べよ。

(2)　累積ハザード関数 $H(t)$ が凸関数のとき寿命分布は IFR であり，それが凹関数のとき分布は DFR であることを示せ。なお，関数 $f(t)$ が区間 I で凸関数であるとは，区間内の任意の 2 点 $t_1 < t_2$ および $0 < p < 1$ なる任意の p に対し，

$$f(pt_1 + (1 - p)t_2) \leq pf(t_1) + (1 - p)f(t_2) \qquad (*)$$

が成り立つことをいい，$f(t)$ が凹関数であるとは式 $(*)$ の不等号 $(\leq)$ が逆向きの $\geq$ となることをいう。

〔4〕　パラメータ1の指数分布に従う確率変数をXとする。Xの累積分布関数は$F(x) = P(X \leq x) = 1 - e^{-x}$である。$\beta$を正の実数とし，$T = X^{1/\beta} = \sqrt[\beta]{X}$と変数変換する。

(1)　Tの確率密度関数$g_\beta(t)$およびハザード関数$h_\beta(t)$を求め，βの値とTの分布のIFR性およびDFR性との関係を示せ。

(2)　$\beta = \dfrac{1}{2}$および$\beta = 2$としたときのTのハザード関数$h_{\frac{1}{2}}(t), h_2(t)$を求め，$0 < t < 5$でそれぞれの関数の概形を図示せよ。

問 2　ある部品の熱処理工程は，特性 A を管理特性とし，$\bar{X}-R$ 管理図（平均値 $\bar{X}$ と範囲 R の 2 つの管理図）で管理されている。図 1 の $\bar{X}-R$ 管理図に示すように，$\bar{X}-R$ 管理図の群の大きさは 4 であり，群は熱処理バッチに対応する。また，図 2 は半導体ウェハのある製造工程におけるウェハ上に付着するパーティクル（微少なごみ）の数を管理特性とした C 管理図である。

管理図は中心線（Center Line）と管理限界線（Control Limit）から成る。管理図の Center Line（CL）は，管理図に打点される統計量の平均値であり，CL の上側に UCL（Upper Control Limit），下側に LCL（Lower Control Limit）が設定され，それらの値は，

$$\mathrm{CL} \pm (\text{管理図に打点される統計量の標準偏差}) \times 3$$

である。$\bar{X}-R$ 管理図での上式における標準偏差は，R 管理図によって安定していることを判断した群内変動より求める。標準偏差の 3 倍で管理限界を設定する方法を 3 シグマルールという。ただし，R 管理図は LCL が負になる場合は考えない。$\bar{X}-R$ 管理図は正規分布を仮定して作成される。

また，C 管理図はポアソン分布を仮定して作成される。図 2 の CL は平均値 1.62 であり，管理限界線は 3 シグマルールで計算している。ただし，R 管理図と同様に，LCL が負になる場合は考えない。

図 1 の $\bar{X}$ 管理図および図 2 の C 管理図ともに管理限界線を越えた点が多発している。この現象に関連して以下の各問に答えよ。

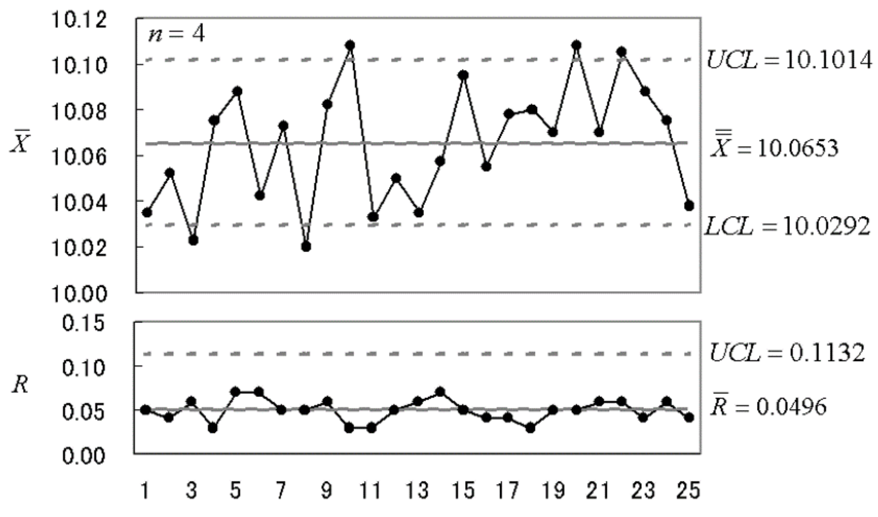


図 1: 熱処理工程の $\bar{X}-R$ 管理図

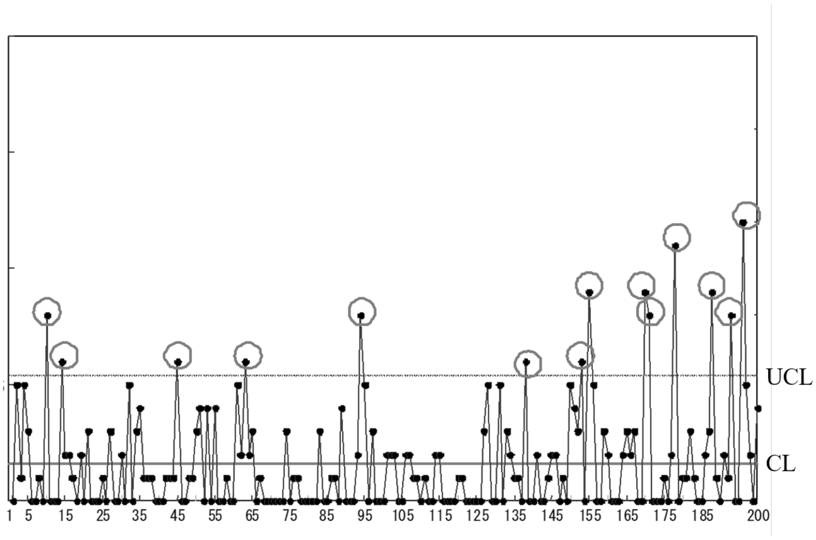


図 2: 半導体ウェハの製造工程の C 管理図

〔1〕 R 管理図から，熱処理のバッチ内（群内）変動は安定しているとみなされる。図 1 の $\bar{X}$ の管理図より，処理バッチ内変動（標準偏差 $\hat{\sigma}_W$）を推定せよ。

〔2〕 $\bar{X}$ 管理図は多数の点が管理限界線を外れている。図 1 の元データ x_{ij} $(i = 1, \ldots, 25;$ $j = 1, 2, 3, 4)$ の標準偏差 $\hat{\sigma}_{\text{Total}}$ が 0.0481 であったとしたとき，熱処理バッチ間変動（群間変動 $\hat{\sigma}_B$）を標準偏差で求めよ。

〔3〕 上問〔2〕のバッチ間変動のほとんどが鋼材の変動によることが判明した。ただし，鋼材の変動は鋼材メーカーの問題であり，その変動は熱処理工程からすれば管理外である。鋼材ロットは処理バッチに対応している。熱処理担当のエンジニアは，熱処理後の特性 A の現状のばらつきは通常のばらつきであり，規格に対しても何ら問題はなく，現状を維持管理したいと判断した。3 シグマルールによって管理するとしたとき，$\bar{X} - R$ 管理図に対する対応策を考えよ。

〔4〕 図 2 の C 管理図は 3 シグマルールを用いている。C 管理図の UCL を求めよ。

〔5〕 図 2 の C 管理図は管理限界外の点が多発している。上問〔3〕と同様に，担当のエンジニアは，この程度のばらつきは通常のばらつきであり問題はなく，現状を維持管理したいと考えている。管理限界外の点が多発している図 2 の現象を説明し，C 管理図を活用するための対応策を考えよ。

問 3　ある窯を用いたタイル焼成工程では，焼成後のタイル強度 y を改善するために次の 4 因子による実験を行い，最も強度が高くなるタイル焼成温度，配合などを求める。

A：タイル焼成温度　B：成分 B 配合量
C：成分 C 配合量　　D：焼成炉内位置

すべての因子は 2 水準である。実験の計画には $L_8(2^7)$ 型直交表を用いる。その直交表を表 1 の左半分に示す。表 1 の右半分は実験を行う順序に関する 4 種類の候補である。

表 1: 直交表と実験の順序候補

No.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	順序 1	順序 2	順序 3	順序 4
1	1	1	1	1	1	1	1	8	5	6	3
2	1	1	1	2	2	2	2	7	3	8	4
3	1	2	2	1	1	2	2	6	6	7	8
4	1	2	2	2	2	1	1	5	1	5	7
5	2	1	2	1	2	1	2	4	8	2	1
6	2	1	2	2	1	2	1	3	4	1	2
7	2	2	1	1	2	2	1	2	7	4	5
8	2	2	1	2	1	1	2	1	2	3	6
成分記号	a		a		a		a				
		b	b			b	b				
				c	c	c	c				
	1 群	2 群		3 群							

〔1〕　因子 A，B，C，D の主効果をすべて推定するため，次の「因子」の行のように各因子を割り付ける。

No.	[1]	[2]	[3]	[4]	[5]	[6]	[7]
因子	A	B		C			D
平方和	5	4	2	1	1	3	1

この表の「平方和」の行には実験結果 y_i $(i = 1, \ldots, 8)$ から求めた列ごとの平方和をあわせて示している。平方和は，第 $[k]$ 列の記号が 1 のときの平均 $\bar{y}_{[k]1}$ と記号が 2 のときの平均 $\bar{y}_{[k]2}$ を用いて $2(\bar{y}_{[k]1} - \bar{y}_{[k]2})^2$ で求められる。因子 A，B，C，D の主効果以外の変動は誤差とみなし，因子 A，B の主効果について F 値を計算せよ。また，完全無作為化実験の実験順序の例として適切なのは，表 1 の右半分の順序 1 から順序 4 のうちのどれかを示せ。

〔2〕　上問〔1〕で，因子 A，B，C，D の主効果に加え，交互作用 A×B もモデルに取り込む。このとき，それら以外の変動を誤差とみなして交互作用 A×B の F 値を計算せよ。

〔3〕　焼成窯ではいくつかのタイルを同時に焼成することができる。そこで，因子 A を 1 次因子，因子 B，C，D を 2 次因子とした分割実験とし，次の表の「因子」の行のような割り付けを行う。なお，1 次単位は {No. 1, No. 2}，{No. 3, No. 4}，{No. 5, No. 6}，{No. 7, No. 8} の 4 つとする。

No.	[1]	[2]	[3]	[4]	[5]	[6]	[7]
因子		A		B	C		D

この場合の実験順序の例として適切なものを，表 1 の順序 1 から順序 4 から選べ。

〔4〕　上問〔1〕の完全無作為化実験の場合，および上問〔3〕の分割実験の場合のそれぞれについて，タイル焼成は何回行うのかを述べよ。

〔5〕　分割実験を実施した際の，上問〔1〕と同様に求めた平方和を次の表の「平方和」の行に示す。1 次誤差，2 次誤差ともに無視できないものとして，因子 A，B，C，D の F 値を計算せよ。

No.	[1]	[2]	[3]	[4]	[5]	[6]	[7]
因子		A		B	C		D
平方和	2	5	2	4	1	1	1

問 4　時系列データ $(\ldots, X_{-1}, X_0, X_1, \ldots, X_n, \ldots)$ は 1 次の自己回帰（AR(1)）モデル

$$X_t = \phi X_{t-1} + \epsilon_t \quad (t = \ldots, -1, 0, 1, \ldots) \tag{1}$$

に従うとする。ここで ϵ_t は互いに独立に $N(0, \sigma^2)$ に従う確率変動項であり，定常性の条件 $|\phi| < 1$ を仮定する。$(X_1, \ldots, X_n)$ につき以下の各問に答えよ。

〔1〕　モデル (1) における $(X_1, \ldots, X_n)$ の自己共分散行列 $T = \{\tau_{ij}\}$ の各成分は

$$\tau_{ij} = \frac{\sigma^2}{1-\phi^2}\phi^{|i-j|} \quad (i, j = 1, \ldots, n)$$

で与えられ，自己相関行列は $R = \{\rho_{ij}\} = \{\phi^{|i-j|}\}$ となることを示せ。

〔2〕　n 次対称行列 $A = \{a_{ij}\}$ を

$$a_{ij} = \begin{cases} 1 & (i = j = 1, i = j = n) \\ 1+\phi^2 & (i = j = 2, \ldots, n-1) \\ -\phi & (|i-j| = 1) \\ 0 & (|i-j| \geq 2) \end{cases}$$

とする。たとえば $n = 4$ では

$$A = \begin{pmatrix} 1 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{pmatrix}$$

である。

一般の n および上問〔1〕の行列 T に対し，$\dfrac{1}{\sigma^2}T$ の逆行列は A で与えられることを示せ。また，A の行列式 $|A|$ および R の行列式 $|R|$ の値を求めよ。

〔3〕　$\boldsymbol{x} = (x_1, \ldots, x_n)'$ を n 次ベクトルとしたとき（プライム $(')$ は転置を表す），上問〔2〕の行列 A に関する 2 次形式 $Q_A = \boldsymbol{x}'A\boldsymbol{x}$ を $x_1, \ldots, x_n$ を用いて書き下し，$|\phi| < 1$ のとき，Q_A はすべての $\boldsymbol{x} \neq \boldsymbol{0}$（成分がすべて 0 のベクトル）に対して常に正であること，すなわち A は正定値であることを示せ。

〔4〕　上問〔2〕の行列 A の $(1,1)$ 要素の 1 のみを ϕ^2 に変えた行列を B とする。たとえば $n=4$ の場合は

$$B = \begin{pmatrix} \phi^2 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{pmatrix}$$

である。一般の n について，B に関する 2 次形式 $Q_B = \boldsymbol{x}'B\boldsymbol{x}$ はすべての $\boldsymbol{x}$ に対して ϕ の値によらず非負となること，すなわち Q_B は非負定値であることを示せ。また，$Q_B = 0$ となる $\boldsymbol{x}$ (ただし，$\boldsymbol{x} \neq \mathbf{0}$) はどのようなベクトルであるか。

〔5〕　モデル (1) における自己回帰係数 ϕ が既知もしくはきわめて精度良く推定されているが誤差分散 σ^2 は未知であるとき，σ^2 の 95 %信頼区間の構成法を示せ。

問 5　ABO 血液型の分布は O 型，A 型，B 型，AB 型の比率で示され，この比率は国や地域によって違いが見られる。日本のある地域 C から無作為に抽出した 100 人を調べたところ，血液型の分布は表 1 のようになった。以下の各問に答えよ。

表 1: 地域 C の血液型分布（観測度数）

血液型	O 型	A 型	B 型	AB 型	合計
観測度数	24	48	16	12	100

〔1〕　日本人の ABO 血液型は

$$\text{O 型} : \text{A 型} : \text{B 型} : \text{AB 型} = 3 : 4 : 2 : 1 \tag{1}$$

の比率で分布すると言われている。帰無仮説を 式 (1) の比率とし，表 1 について適合度のカイ 2 乗検定を有意水準 5%で行い，その結果を述べよ。

〔2〕　血液型の分布の観測度数が，k を自然数として $6k$，$12k$，$4k$，$3k$ であったとしたとき（表 1 では $k = 4$），適合度のカイ 2 乗検定が有意水準 5%で有意になる最小の k はいくらか。

〔3〕　一般に，適合度のカイ 2 乗検定統計量は，近似的にカイ 2 乗分布に従うことからその名があるが，その統計量が近似的にカイ 2 乗分布に従う根拠は何かを詳細に述べよ。厳密に証明する必要はない。

問題は次ページに続く。

〔4〕　ABO 血液型は，親から受け継いだ 3 つの遺伝子 O，A，B の組合せによって決まることが知られていて，表 2 のように血液型が決まる。これより，遺伝子 O，A，B はそれぞれ r，p，q $(r+p+q=1)$ の比率で分布しているとすると，各血液型の比率は表 2 の最後の行に示したようになる。

表 2: 遺伝子を考慮した血液型分布

血液型	O 型	A 型	B 型	AB 型
遺伝子型	OO	AA AO OA	BB BO OB	AB BA
比率	r^2	p^2+2pr	q^2+2qr	$2pq$

全観測度数を N とし，各血液型の観測度数をそれぞれ n_O，n_A，n_B，n_{AB}，各遺伝子型の度数を f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} とする（f_{AO}，f_{BO}，f_{AB} はそれぞれ AO と OA，BO と OB，AB と BA の合計度数である）。このとき，$f_{OO} = n_O$，$f_{AA} + f_{AO} = n_A$，$f_{BB} + f_{BO} = n_B$，$f_{AB} = n_{AB}$ であり，f_{AA}，f_{AO}，f_{BB}，f_{BO} は実際は観測されない度数である。

比率 r，p，q の最尤推定値を求める。度数 f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} に基づく尤度関数は，r，p，q に依存しない定数を無視すると

$$L(r,\, p,\, q) \propto (r^2)^{f_{OO}}(p^2)^{f_{AA}}(2pr)^{f_{AO}}(q^2)^{f_{BB}}(2qr)^{f_{BO}}(2pq)^{f_{AB}}$$

となる。次の (i) および (ii) に答え，最尤推定値を求める数値計算の反復法を構築せよ。ただし実際に数値を求める必要はない。

(i)　ラグランジュの未定乗数を λ とした

$$Q = \log L(r,\, p,\, q) - \lambda(r+p+q-1)$$

を r，p，q でそれぞれ偏微分して 0 と置き，$L(r,\, p,\, q)$ を最大化する r，p，q の値を求める式を示せ。

(ii)　上記 (i) で求めた r，p，q を用いて度数 f_{AA}，f_{AO}，f_{BO}，f_{BB} の期待値を求める式を示せ。

医薬生物学

問 1　T を生存時間を表す確率変数とし，n 人について観測された生存時間を $t_1, t_2, \ldots, t_n$ とする。$r \leq n$ となる r 個のイベントの観察時点を $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ となるように昇順に並べ替え，j 番目を $t_{(j)}$ $(j = 1, 2, \ldots, r)$ と表す。時点 $t_{(j)}$ の直前にイベントが起きる可能性のある人数（リスク集合）を n_j，時点 $t_{(j)}$ におけるイベント数を d_j とする。5 名を観察した結果，表 1 のような生存時間データが得られたとする。

表 1: 生存時間データ

生存時間（週）	打ち切りの有無 （1:打ち切り，0:イベント）
10	0
13	0
18	1
19	0
23	1

このとき，以下の各問に答えよ。

〔1〕　表 1 のデータから表 2 を作成せよ。さらに，カプラン・マイヤー法による生存曲線を図示せよ。ただし，$\hat{S}(t)$ は時点 t におけるカプラン・マイヤー法により推定された生存関数の推定値とする。

表 2: 生存関数のカプラン・マイヤー推定値

生存時間（週）	n_j	d_j	$\hat{S}(t)$
10			
13			
19			

〔2〕　$t_{(j)} \leq t < t_{(j+1)}$ $(j = 1, 2, \ldots, r)$ における生存関数のネルソン・アーレン推定値は次式で与えられる。

$$\tilde{S}(t) = \prod_{k=1}^{j} \exp\left(-\frac{d_k}{n_k}\right)$$

任意の時点 t において $\tilde{S}(t) \geq \hat{S}(t)$ であることを示せ。

〔3〕　表 1 のデータのように観察されたデータのうち最長の生存時間データが打ち切りの場合，カプラン・マイヤー法により推定された生存曲線は 0 に到達しない。そのため，生存曲線の曲線下面積が定義できず，平均生存時間を推定することはできない。このような場合に，平均生存時間の代替指標として，境界内平均生存時間を用いることがある。境界内平均生存時間は，境界時間 τ 内での生存時間を $X(\tau)$ $(= \min(T, \tau))$ としたとき，$X(\tau)$ の期待値として定義される。境界内平均生存時間は，境界時間 τ 内における生存曲線の曲線下面積に等しくなることを示せ。

〔4〕　生存時間 $X(\tau)$ の分散 $\mathrm{Var}[X(\tau)]$ を導出せよ。ただし，生存関数を $S(t) = \exp(-\lambda t)$ とする。

〔5〕　生存関数を $S(t) = \exp(-\lambda t)$ とする。表 1 のデータからハザード λ の最尤推定値を求めよ。さらに，境界時間を $\tau = 20$ とした境界内平均生存時間とその分散 $\mathrm{Var}[X(\tau)]$ の推定値を求めよ。ただし，指数関数の値は p.56 の付表 5 を参照すること。

問 2　乳がんの切除後に化学療法を受けた女性グループ ($X = 1$) と受けなかった女性グループ ($X = 0$) を 5 年間追跡した仮想的なコホート研究の結果を以下に示す（*Epidemiology* 2019; 30: 541–548）。共変量 $\boldsymbol{Z} = (\mathrm{Memo}, \mathrm{Grade})$ は，閉経の有無（1: あり，0: なし）とがんの進行度（1: 3 以上，0: 1 または 2）を表し，いずれも再発 D（1: 再発あり，0: なし）のリスク因子として知られている。以下の各問に答えよ。

	$\boldsymbol{Z}$		$X = 1$		$X = 0$	
層 k	Memo	Grade	N	$D = 1$	N	$D = 1$
1	0	0	172	12	1028	58
2	0	1	220	65	180	45
3	1	0	209	48	171	35
4	1	1	597	353	103	57
Total			$N_1 = 1198$	$Y_1 = 478$	$N_0 = 1482$	$Y_0 = 195$

Memo: 閉経の有無，Grade: がんの進行度

〔1〕　$X = x$ のグループの人数を N_x，再発数を Y_x とし，Y_x は再発確率（リスク）p_x の二項分布に従うとする。データを共変量で層別せずに，未調整のリスク差 $p_1 - p_0$ の推定値とその 95%信頼区間を有効数字 2 桁で求めよ。ただし，p_x の最尤推定量 $\hat{p}_x = \dfrac{Y_x}{N_x}$ とその標準誤差の推定量 $\left\{\dfrac{Y_x(N_x - Y_x)}{N_x^3}\right\}^{1/2}$ を用いて正規近似を行うこととする。

〔2〕　共変量 $\boldsymbol{Z}$ で層別したデータについて，層 $k\ (= 1, \ldots, 4)$ のグループ $X = x$（人数 N_{xk}）の再発数を Y_{xk} とする。さらに，各層の人数の割合を $w_k = \dfrac{N_{0k} + N_{1k}}{N_0 + N_1}$，$\displaystyle\sum_{k=1}^{4} w_k = 1$ とし，各層の $X = x$ の再発確率（リスク）p_{xk} の重み付き平均 $R_{xw} = \displaystyle\sum_{k=1}^{4} w_k p_{xk}$ を調整リスクとする。調整リスク差 $R_{1w} - R_{0w}$ の推定値を有効数字 2 桁で求めよ。

〔3〕　各層内での治療確率 $e(\boldsymbol{z}) = P(X = 1|\boldsymbol{Z} = \boldsymbol{z})$ を傾向スコアという。傾向スコアの値が同じ集団では，層別に用いた共変量 $\boldsymbol{Z}$ の同時分布が治療群間で等しくなる (傾向スコアのバランス特性)。傾向スコアを求めるために，X を結果変数としてロジスティック回帰モデルを最尤法で当てはめたところ，以下の出力を得た。

Parameter	Estimate	Standard Error	P
Intercept	-1.7879	0.0824	< 0.0001
Memo	1.9885	0.1320	< 0.0001
Grade	1.9885	0.1300	< 0.0001
Memo$*$Grade	-0.4320	0.1972	0.0285

これらの結果から，各層の傾向スコアを推定できる。推定された傾向スコアのバランス特性をデータから示せ。

〔4〕　上問〔3〕から得られた傾向スコアの値で層別したデータに対して，上問〔2〕と同様の推定量によりグループ $X = 1$ と $X = 0$ の調整リスク差の推定値を有効数字 2 桁で求めよ。

〔5〕　上問〔2〕の傾向スコアは次式のように表現できる。

$$e(\boldsymbol{z}) = P(X = 1|\boldsymbol{Z} = \boldsymbol{z}) = E[X|\boldsymbol{Z} = \boldsymbol{z}]$$

このとき，次式の傾向スコアのバランス特性が成り立つことを証明せよ。

$$P(X = 1|\boldsymbol{Z} = \boldsymbol{z}, e(\boldsymbol{Z}) = e(\boldsymbol{z})) = P(X = 1|e(\boldsymbol{Z}) = e(\boldsymbol{z}))$$

ただし，$P(X = 1|\boldsymbol{Z} = \boldsymbol{z}, e(\boldsymbol{Z}) = e(\boldsymbol{z})) = P(X = 1|\boldsymbol{Z} = \boldsymbol{z})$ が成り立つことを用いてよい。

問 3　ある疾患 D に罹患しているか否かは生検によって確定診断されるが，このためのスクリーニング検査として検査法 A がある。今，この検査法 A とは別に，新たな検査法 B が開発された。このとき，検査法 A と検査法 B の診断性能を比較したい。そこで，被験者 200 名全員に検査法 A と検査法 B，および生検を受けてもらい，その試験の結果を，生検の結果に基づき実際に疾患 D に罹患していると判明した群と疾患 D に罹患していないと判明した群とに分けてまとめたものが次の表である。このとき以下の各問に答えよ。

疾患 D に罹患している群

		検査法 B		
		陽性	陰性	計
検査法	陽性	65	7	72
A	陰性	16	2	18
計		81	9	90

疾患 D に罹患していない群

		検査法 B		
		陽性	陰性	計
検査法	陽性	2	16	18
A	陰性	7	85	92
計		9	101	110

〔1〕　この試験のデータから推定される検査法 A および検査法 B の感度を求めよ。また，同様に検査法 A および検査法 B の陽性的中率を求めよ。

〔2〕　検査法 A と B の真の感度を比較したい。このとき次の仮説

$$\begin{cases} H_0 : \text{検査法 A の真の感度} = \text{検査法 B の真の感度} \\ H_1 : \text{検査法 A の真の感度} \neq \text{検査法 B の真の感度} \end{cases}$$

に対して，正規近似を用いたスコア型の検定統計量（ただし連続性の補正は行わなくてよい）により有意水準 5%で検定せよ。
（ヒント：疾患 D に罹患している群のみを考えればよく，(検査法 A，検査法 B）の結果が異なるという条件のもとで，(陽性，陰性）セルの観測度数は，帰無仮説 H_0 のもとで $\mathrm{Bin}(23, 0.5)$ の二項分布に従うことを用いる。）

〔3〕　検査法 A と検査法 B の真の陽性的中率に関する仮説

$$\begin{cases} H_0 : \text{検査法 A の真の陽性的中率} = \text{検査法 B の真の陽性的中率} \\ H_1 : \text{検査法 A の真の陽性的中率} \neq \text{検査法 B の真の陽性的中率} \end{cases}$$

を検定することを考える。
上記の陽性的中率に関する仮説検定を考えるにあたって，先の観測度数に基づく表を改めて次のように示すことにする。

疾患 D に罹患している群

		検査法 B		
		陽性	陰性	計
検査法	陽性	x_{++D}	x_{+-D}	$x_{+\bullet D}$
A	陰性	x_{-+D}	x_{--D}	$x_{-\bullet D}$
計		$x_{\bullet+D}$	$x_{\bullet-D}$	$x_{\bullet\bullet D}$

疾患 D に罹患していない群

		検査法 B		
		陽性	陰性	計
検査法	陽性	$x_{++\overline{D}}$	$x_{+-\overline{D}}$	$x_{+\bullet\overline{D}}$
A	陰性	$x_{-+\overline{D}}$	$x_{--\overline{D}}$	$x_{-\bullet\overline{D}}$
計		$x_{\bullet+\overline{D}}$	$x_{\bullet-\overline{D}}$	$x_{\bullet\bullet\overline{D}}$

また，上の表に対する真の確率構造を

疾患 D に罹患している群

		検査法 B		
		陽性	陰性	計
検査法	陽性	π_{++D}	π_{+-D}	$\pi_{+\bullet D}$
A	陰性	π_{-+D}	π_{--D}	$\pi_{-\bullet D}$
計		$\pi_{\bullet+D}$	$\pi_{\bullet-D}$	$\pi_{\bullet\bullet D}$

疾患 D に罹患していない群

		検査法 B		
		陽性	陰性	計
検査法	陽性	$\pi_{++\overline{D}}$	$\pi_{+-\overline{D}}$	$\pi_{+\bullet\overline{D}}$
A	陰性	$\pi_{-+\overline{D}}$	$\pi_{--\overline{D}}$	$\pi_{-\bullet\overline{D}}$
計		$\pi_{\bullet+\overline{D}}$	$\pi_{\bullet-\overline{D}}$	$\pi_{\bullet\bullet\overline{D}}$

とする。ここで，$\boldsymbol{x} = (x_{++D}, x_{+-D}, \ldots, x_{-+\overline{D}}, x_{--\overline{D}})'$（記号 $'$ はベクトルの転置を表す）は，確率ベクトル $\boldsymbol{\pi} = (\pi_{++D}, \pi_{+-D}, \ldots, \pi_{-+\overline{D}}, \pi_{--\overline{D}})'$ をもつ多項分布に従うとする。このとき，$x_{\bullet\bullet D} + x_{\bullet\bullet\overline{D}} = n$，$\boldsymbol{p} = \dfrac{\boldsymbol{x}}{n}$ とすると，$n \to \infty$ で

$$\sqrt{n}(\boldsymbol{p} - \boldsymbol{\pi})$$

は多変量中心極限定理により近似的に平均ベクトルが $\mathbf{0}$ の多変量正規分布に従う。この統計量の分散共分散行列を $\boldsymbol{\pi}$ またはその成分を用いて表せ。

〔4〕

$$f(\boldsymbol{\pi}) = \frac{\pi_{+\bullet D}}{\pi_{+\bullet D} + \pi_{+\bullet\overline{D}}} - \frac{\pi_{\bullet+D}}{\pi_{\bullet+D} + \pi_{\bullet+\overline{D}}}$$

とするとき，デルタ法の近似による $\sqrt{n}f(\boldsymbol{p})$ の分散を〔3〕で用いた行列によって表現せよ（要素まで計算する必要はない）。

〔5〕　上問〔3〕および〔4〕を用いて,〔3〕で与えられた陽性的中率に関する仮説を有意水準 5%で検定せよ。ただし,〔4〕で求めた分散における未知の確率 $\boldsymbol{\pi}$ を $\boldsymbol{p}$ に置き換えた分散の推定値は 0.42 であることを用いてよい。

問 4　ある希少疾患の臨床試験において，計 6 名の被験者にランダムに薬剤 A もしくは薬剤 B を投与し，ある検査の観察終了時の測定値を表にまとめた。以下の各問に答えよ。

群	測定値			平均	標準偏差
薬剤 A	2.79	4.64	7.03	4.82	2.13
薬剤 B	0.21	0.73	5.52	2.15	2.93

〔1〕　上表のデータに対して，両側有意水準 5%で帰無仮説を「2 つの群における母平均は等しい」とする Student の t 検定を適用し，有意性を判定せよ。また，この検定の適用が妥当であるための条件を 3 つ示せ。

〔2〕　上表のデータに対して，ノンパラメトリック法である Wilcoxon の順位和検定を適用する。この検定における帰無仮説を述べよ。

〔3〕　上問〔2〕で適用する Wilcoxon の順位和検定における検定統計量を示し，上表のデータにおけるその検定統計量の実現値を求めよ。

〔4〕　上問〔3〕で示された Wilcoxon の検定統計量に対して，帰無仮説のもとでの取り得る値とそれぞれの値を取る確率を求めよ。

〔5〕　上表のデータに対して，Wilcoxon の順位和検定を適用したときの正確な両側 p 値を求めよ。さらに，このデータに対して上記の検定を有意水準 5%で行うときの第 1 種の過誤確率を求め，この状況における Wilcoxon の順位和検定の問題点を指摘せよ。

問 5　ABO 血液型の分布は O 型，A 型，B 型，AB 型の比率で示され，この比率は国や地域によって違いが見られる。日本のある地域 C から無作為に抽出した 100 人を調べたところ，血液型の分布は表 1 のようになった。以下の各問に答えよ。

表 1: 地域 C の血液型分布（観測度数）

血液型	O 型	A 型	B 型	AB 型	合計
観測度数	24	48	16	12	100

〔1〕　日本人の ABO 血液型は

$$\text{O 型}:\text{A 型}:\text{B 型}:\text{AB 型} = 3:4:2:1 \tag{1}$$

の比率で分布すると言われている。帰無仮説を 式 (1) の比率とし，表 1 について適合度のカイ 2 乗検定を有意水準 5%で行い，その結果を述べよ。

〔2〕　血液型の分布の観測度数が，k を自然数として $6k$，$12k$，$4k$，$3k$ であったとしたとき（表 1 では $k = 4$），適合度のカイ 2 乗検定が有意水準 5%で有意になる最小の k はいくらか。

〔3〕　一般に，適合度のカイ 2 乗検定統計量は，近似的にカイ 2 乗分布に従うことからその名があるが，その統計量が近似的にカイ 2 乗分布に従う根拠は何かを詳細に述べよ。厳密に証明する必要はない。

問題は次ページに続く。

〔4〕 ABO 血液型は，親から受け継いだ 3 つの遺伝子 O，A，B の組合せによって決まることが知られていて，表 2 のように血液型が決まる。これより，遺伝子 O，A，B はそれぞれ r，p，q $(r+p+q=1)$ の比率で分布しているとすると，各血液型の比率は表 2 の最後の行に示したようになる。

表 2: 遺伝子を考慮した血液型分布

血液型	O 型	A 型	B 型	AB 型
遺伝子型	OO	AA AO OA	BB BO OB	AB BA
比率	r^2	p^2+2pr	q^2+2qr	$2pq$

全観測度数を N とし，各血液型の観測度数をそれぞれ n_O，n_A，n_B，n_{AB}，各遺伝子型の度数を f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} とする（f_{AO}，f_{BO}，f_{AB} はそれぞれ AO と OA，BO と OB，AB と BA の合計度数である）。このとき，$f_{OO}=n_O$，$f_{AA}+f_{AO}=n_A$，$f_{BB}+f_{BO}=n_B$，$f_{AB}=n_{AB}$ であり，f_{AA}，f_{AO}，f_{BB}，f_{BO} は実際は観測されない度数である。

比率 r，p，q の最尤推定値を求める。度数 f_{OO}，f_{AA}，f_{AO}，f_{BB}，f_{BO}，f_{AB} に基づく尤度関数は，r，p，q に依存しない定数を無視すると

$$L(r,p,q) \propto (r^2)^{f_{OO}}(p^2)^{f_{AA}}(2pr)^{f_{AO}}(q^2)^{f_{BB}}(2qr)^{f_{BO}}(2pq)^{f_{AB}}$$

となる。次の (i) および (ii) に答え，最尤推定値を求める数値計算の反復法を構築せよ。ただし実際に数値を求める必要はない。

(i) ラグランジュの未定乗数を λ とした

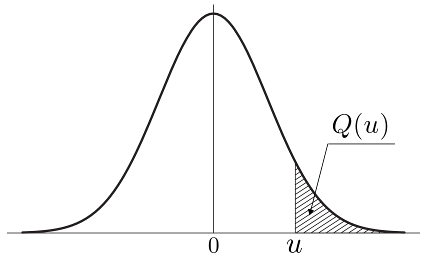
$$Q=\log L(r,p,q)-\lambda(r+p+q-1)$$

を r，p，q でそれぞれ偏微分して 0 と置き，$L(r,p,q)$ を最大化する r，p，q の値を求める式を示せ。

(ii) 上記 (i) で求めた r，p，q を用いて度数 f_{AA}，f_{AO}，f_{BO}，f_{BB} の期待値を求める式を示せ。

付　表

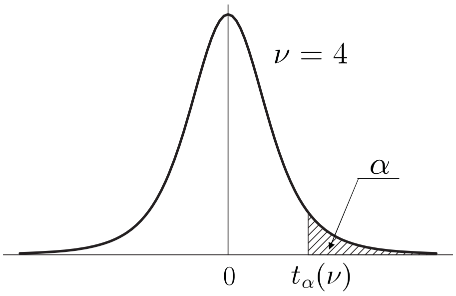
付表 1. 標準正規分布の上側確率



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。

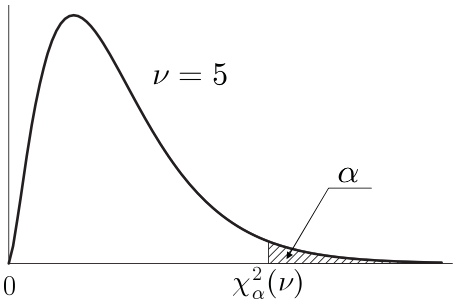
付表 2. t 分布のパーセント点



	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。

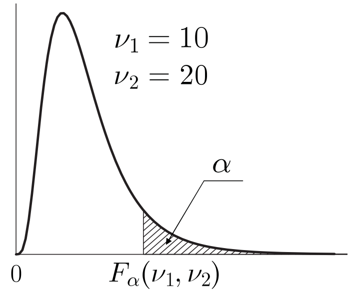
付表 3. カイ二乗分布のパーセント点



	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。

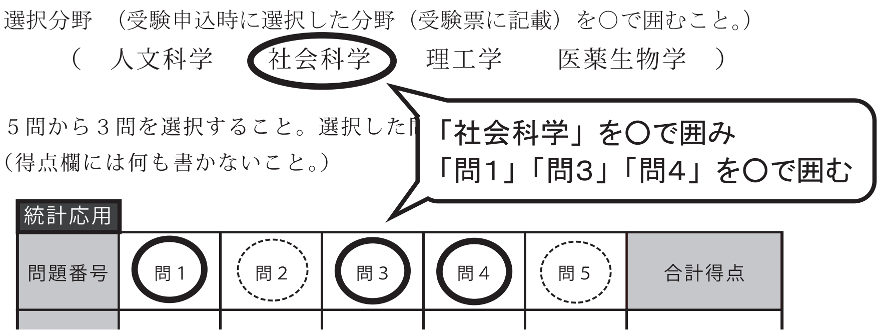
【解答冊子記入例】

- 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**

2019.11