(2014 年 12 月 26 日版)

統計検定準１級例題集　No.1　(問題及び略解)

注意事項：

1. この例題集は 2015 年 6 月開始の統計検定準 1 級の例題集です。
2. 統計検定準 1 級の問題は，I. 選択問題及び部分記述問題，II. 論述問題，の 2 つの部分からなります。
3. 論述問題は 3 題出題され，受験者は 1 題を選択して解答します。
4. それぞれの部分記述問題は，この例題集において **記述 4** のように記載されています。解答は解答用紙の指定されたスペースに記入します。この例題集の最後の解答用紙の例を参考にしてください。
5. それぞれの選択問題は，統計検定 2 級等と同様に **10** のように記載されています。
6. 実際の問題冊子には統計数値表が掲載されます。
7. 論述問題の配点は全体の 3 割程度です。
8. 準 1 級の合格水準は，この例題集において 6 割程度を想定しています。

I. 選択問題及び部分記述問題

問 1　A, B, C の 3 つの事象について次のように確率が与えられている。

$$P(A) = 0.45, P(B) = 0.45, P(C) = 0.4, P(A \cap B) = 0.2,$$
$$P(B \cap C) = 0.1, P(A \cap C) = 0.15, P(A \cap B \cap C) = 0.05$$

〔1〕$P(A \cup B)$ として正しいものを次の ①～⑤ のうちから一つ選べ。 $\boxed{\mathbf{1}}$

① 0.5　② 0.6　③ 0.7　④ 0.8　⑤ 0.9

〔2〕$P(A \cup B \cup C)$ として正しいものを次の ①～⑤ のうちから一つ選べ。 $\boxed{\mathbf{2}}$

① 0.8　② 0.85　③ 0.9　④ 0.95　⑤ 1.0

問 2　次の表はある地域における 1 日の死亡者数の集計結果（500 日間）である。

死亡者数（人）	0	1	2	3	4	5	6 以上	計
件数（日数）	55	144	140	95	45	15	6	500

〔1〕1 日の死亡者数 X がパラメータ λ のポアソン分布にしたがうと仮定するとき，ある日の死亡者数が 3 人である確率を求める式として正しいものはどれか。次の ①～⑤ のうちから適切なものを一つ選べ。 $\boxed{\mathbf{3}}$

① λ　② $\lambda(1-\lambda)^3$　③ ${}_6C_3\lambda^3(1-\lambda)^3$

④ $\lambda e^{-3\lambda}$　⑤ $\dfrac{\lambda^3 e^{-\lambda}}{3!}$

〔2〕1 日の死亡者数 X がパラメータ λ のポアソン分布にしたがうと仮定するとき，X^2 の期待値 $E(X^2)$ とパラメータ λ の関係として正しいものはどれか。次の ①～⑤ のうちから適切なものを一つ選べ。 $\boxed{\mathbf{4}}$

① $E(X^2) = \dfrac{2}{\lambda^2}$　② $E(X^2) = \dfrac{1}{\lambda}$　③ $E(X^2) = \lambda$

④ $E(X^2) = \lambda^2 + \lambda$　⑤ $E(X^2) = \lambda^2$

〔3〕1 日の死亡者数 X がパラメータ λ のポアソン分布にしたがっているか否かの検定を行う。データの平均値をもとにパラメータを推定し，次の表の期待度数を得た。

死亡者数（人）	0	1	2	3	4	5	6 以上
期待度数（日）	67.7	135.3	135.3	90.2	45.1	18.0	8.3

このとき適合度検定の判断として，次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 $\boxed{\textbf{5}}$

① 検定統計量の χ^2 値 =0.647 が自由度 7 の χ^2 分布の上側 5% 点より大きいか比べる。

② 検定統計量の χ^2 値 =0.647 が自由度 6 の χ^2 分布の上側 5% 点より大きいか比べる。

③ 検定統計量の χ^2 値 =4.498 が自由度 7 の χ^2 分布の上側 5% 点より大きいか比べる。

④ 検定統計量の χ^2 値 =4.498 が自由度 6 の χ^2 分布の上側 5% 点より大きいか比べる。

⑤ 検定統計量の χ^2 値 =4.498 が自由度 5 の χ^2 分布の上側 5% 点より大きいか比べる。

問 3 $\begin{pmatrix} X \\ Y \end{pmatrix}$ を次の 2 変量正規分布に従う確率ベクトルとする。

$$N\left(\begin{pmatrix} 1.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 3.0 & 2.0 \\ 2.0 & 4.0 \end{pmatrix}\right)$$

〔1〕 $\begin{pmatrix} X+Y \\ X-Y \end{pmatrix}$ が従う 2 変量正規分布を次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{6}}$

① $N\left(\begin{pmatrix} 3.0 \\ 1.0 \end{pmatrix}, \begin{pmatrix} 9.0 & 0.0 \\ 0.0 & 9.0 \end{pmatrix}\right)$

② $N\left(\begin{pmatrix} 3.0 \\ 1.0 \end{pmatrix}, \begin{pmatrix} 11.0 & 0.0 \\ 0.0 & 9.0 \end{pmatrix}\right)$

③ $N\left(\begin{pmatrix} 3.0 \\ -1.0 \end{pmatrix}, \begin{pmatrix} 11.0 & -1.0 \\ -1.0 & 3.0 \end{pmatrix}\right)$

④ $N\left(\begin{pmatrix} 3.0 \\ -1.0 \end{pmatrix}, \begin{pmatrix} 11.0 & -2.0 \\ -2.0 & 4.0 \end{pmatrix}\right)$

⑤ $N\left(\begin{pmatrix} 3.0 \\ -1.0 \end{pmatrix}, \begin{pmatrix} 11.0 & 0.0 \\ 0.0 & 3.0 \end{pmatrix}\right)$

〔2〕 X を与えた時の Y の条件つき分布を次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{7}}$

① $N(2.0, 4.0)$　② $N(2.0+X, 3.0)$　③ $N(1.33+X, 3.5)$
④ $N(1.33+0.67X, 2.67)$　⑤ $N(1.33+0.67X, 2.5)$

問 4　2013 年 6 月の NHK による政治意識月例調査に回答した 1008 人の内閣支持率は 62 ％であった。

〔1〕母集団の内閣支持率の 95 ％信頼区間を構成したい。最も適切な信頼区間を次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{8}}$

① $0.62 \pm 1.96\sqrt{0.62 \times 0.38}$
② $0.62 \pm 1.96\sqrt{0.62 \times 0.38/1008}$
③ $0.62 \pm 1.64\sqrt{0.62 \times 0.38}/1008$
④ $0.62 \pm 1.64\sqrt{0.62 \times 0.38}$
⑤ $0.62 \pm 1.64\sqrt{0.62 \times 0.38}/1008$

〔2〕次の記述 I～III は，信頼区間の幅を狭くするための方法に関する記述である。
I. 回答者数を増やす。
II. 回答者の若者の割合を増やす。
III. 信頼係数を大きくする。
これら記述の正誤の組合せとして，適切なものを次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{9}}$

① I のみが正しい。　② II のみが正しい。　③ III のみが正しい。
④ I と II が正しい。　⑤ I と III が正しい。

〔3〕真の内閣支持率が 60 ％であるとき，内閣支持率の推定値が 62 ％を超える確率は，およそいくらか。最も適切な値を，次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{10}}$

① 0.01　② 0.05　③ 0.1　④ 0.15　⑤ 0.2

〔4〕内閣支持率の 95 ％信頼区間の幅が 4 ％以内となるためには標本サイズは何人以上必要か。最も適切な値を，次の ① ～ ⑤ のうちから一つ選べ。 $\boxed{\textbf{11}}$

① 40　② 120　③ 600　④ 2400　⑤ 4800

問 5　ある 2 つのタイプ（A と B）の商品について，年齢層により利用傾向が異なるのではないかという意見が出された。この意見についてデータを収集して検討することとした。2 つの商品のタイプのどちらが好きかについて，2 つの年齢層（20 代と 40 代）に属する計 400 人からの回答を集計した結果，次のような分割表が得られた。

これに関して下の問に答えよ。

年齢層と商品タイプの分割表

年齢層＼タイプ	A	B	計
20 代	130	110	240
40 代	70	90	160
計	200	200	400

〔1〕年齢層と商品タイプの関連について考える場合に，年齢層と商品タイプの選択が独立のとき，次の各セルの期待度数を表した分割表の (ア) と (エ) に入る値はいくつか。

(ア) **記述 1**　(エ) **記述 2**

分割表

年齢層＼タイプ	A	B	計
20 代	(ア)	(イ)	240
40 代	(ウ)	(エ)	160
計	200	200	400

〔2〕年齢層と商品タイプの独立性を検定するために，得られた分割表に対して以下のように有意水準を 5% としてカイ 2 乗検定を行った。次の (オ) と (カ) に入る値はいくらか。

検定統計量の χ^2 値を計算すると χ^2 値は (オ) となった。カイ 2 乗分布表より得られる棄却点 (カ) と比較して χ^2 値のほうが大きいので，年齢層により商品の好みに差があると結論する。

(オ) **記述 3**　(カ) **記述 4**

問 6　次のような 2×3 分割表を考える。行和及び列和を固定したもとで x_{11} のとりうる範囲として正しいものを下の ① ～ ⑤ のうちから一つ選べ。 **12**

x_{11}	x_{12}	x_{13}	12
x_{21}	x_{22}	x_{23}	8
10	2	8	20

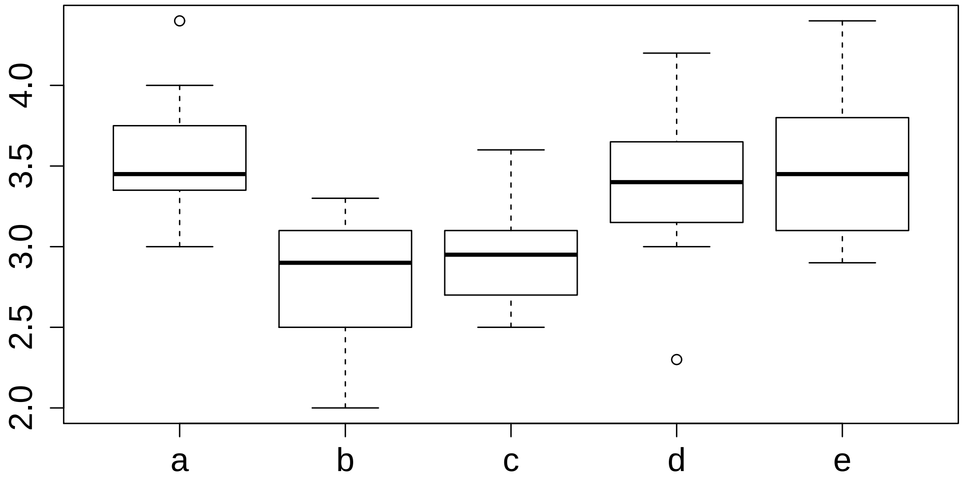
① $0 \leq x_{11} \leq 10$　② $1 \leq x_{11} \leq 10$　③ $2 \leq x_{11} \leq 10$
④ $3 \leq x_{11} \leq 10$　⑤ $4 \leq x_{11} \leq 10$

問 7　同じ植物の 2 つの品種 A と B がある。品種 A は 21 本，品種 B は 16 本についてある部位の長さ (cm) を測定した。下の表は，その平均値と分散である。

	平均値	分散
品種 A	2.78	0.145
品種 B	2.93	0.095

これより，2 つの品種のその部位の長さに差があるかどうかを検定したい。その部位の長さは正規分布にしたがい，分散は等しいと仮定できるものとして t 検定を行う。

〔1〕検定の前に，2 つの標本の分布の様子を箱ひげ図で確認しておく。次の図のうち，品種 A と品種 B はそれぞれどの箱ひげ図にあたるか。下の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **13**



① 品種 A : a，品種 B : e　② 品種 A : b，品種 B : c　③ 品種 A : a，品種 B : d

④ 品種 A : b，品種 B : e　⑤ 品種 A : c，品種 B : d

〔2〕2 つの標本をプールした分散 s^2 を求めよ。 **記述 5**

〔3〕t 統計量の値を求めよ。 **記述 6**

〔4〕この検定の結果として次のような結論を導いた。空欄（ア）～（ウ）に入る言葉として正しい組み合わせはどれか。下の ① ～ ⑤ のうちから適切なものを一つ選べ。 **14**

有意水準 5% で両側検定を用いることとする。自由度 35 の t 分布の上側 2.5% 点は 2.030 である。〔3〕で求めた t 統計量の値から，「2 つの品種のある部位の長さに（ア）」という帰無仮説は（イ）。よって「2 つの品種のある部位の長さに（ウ）」と結論する。

① （ア）差はない　（イ）棄却される　（ウ）差があるといえる
② （ア）差はある　（イ）棄却される　（ウ）差があるといえる
③ （ア）差はある　（イ）棄却されない　（ウ）差があるといえる
④ （ア）差はない　（イ）棄却されない　（ウ）差があるとはいえない
⑤ （ア）差はある　（イ）棄却される　（ウ）差があるとはいえない

問 8　R.A.Fisher の 1936 年の論文にある 3 種（setosa，versicolor，virginica）のあやめの「がく片の長さ」のデータを利用して分析した結果を考察する。このデータでは，それぞれ 50 ずつの個体が観測されている。

〔1〕 3 種の「がく片の長さ」の等分散性について有意水準 5 %の F 検定を行った。次の表は，その出力結果の一部である。num は分子，denom は分母の略語である。

出力結果の一部

・setosa と versicolor の結果（前者が分子，後者が分母，以下同様）
```
F = 0.4663, num df = 49, denom df = 49, p-value = 0.008657
95 percent confidence interval: 0.2646385 0.8217841
```
・versicolor と virginica の結果
```
F = 0.6589, num df = 49, denom df = 49, p-value = 0.1478
95 percent confidence interval: 0.3739257 1.1611546
```
・setosa と virginica の結果
```
F = 0.3073, num df = 49, denom df = 49, p-value = 6.366e-05
95 percent confidence interval: 0.1743776 0.5414962
```

この出力結果に関する説明として，最も適切なものを次の ① ～ ⑤ のうちから一つ選べ。 **15**

① 有意水準 5%で有意でないのは，versicolor と virginica の分散の差異である。
② どの 95%信頼区間も 0 を含んでいないので，分散の差異がみられる種の組み合わせはない。
③ versicolor と virginica の分散については，95%信頼区間が 1 を含んでいるので，有意性を判断できない。
④ F 検定において，信頼区間と p-値に関係はない。
⑤ 一般に，F-値が小さい方が分散の差異が大きいと言える。

〔2〕setosa と versicolor のがく片の長さについて，平均値の差の検定を行った。次の表は，有意水準 5%の Welch の t 検定を用いた出力結果の一部である。

出力結果の一部

```
・setosa と versicolor の結果
  t = -10.521, df = 86.538, p-value  < 2.2e-16
  mean of setosa mean of versicolor
  5.006          5.936
```

Welch の t 検定と出力結果に関する説明として，適切でないものを次の ①～⑤ のうちから一つ選べ。 **16**

① このデータで分散の差異が全くないなら，Welch の t 検定の自由度は 98 である。

② 個体数が同じなら，Welch の t 検定の自由度は分散の差が大きいほど小さくなる。

③ この結果から，Student の t 検定を用いたなら，有意差はないという結果になることがわかる。

④ 個体数が同じなら， Welch の t 検定の t-値と Student の t 検定の t-値は同じ値である。

⑤ 各種類において，個体数が異なる場合でも Welch の t 検定を利用できる。

〔3〕 3 種のあやめの「がく片の長さ」の平均値の差の検定を行うとき，単純に 3 回の t 検定を繰り返して判断してはいけない理由として，最も適切なものを次の ①～⑤ のうちから一つ選べ。 **17**

① 3 回のいずれか一つが棄却される確率が高くなるから

② 3 回の検定を行う順番を変更すると結果が変わるから

③ 互いの分散が違うから

④ 検出力が悪くなるから

⑤ 3 つの種類はたまたま選ばれただけで他にもあるから

問 9 確率変数 W は自由度 m のカイ二乗分布に従うとする。m が大きい時，W 及び $\sqrt{W}$ の分布は正規分布で近似することができる。それぞれの正規分布の適切な組合せを次の ①～⑤ のうちから一つ選べ。 **18**

① $W : N(m, m), \quad \sqrt{W} : N(\sqrt{m}, \sqrt{m})$

② $W : N(m, m), \quad \sqrt{W} : N(\sqrt{m}, 1)$

③ $W : N(m, 2m), \quad \sqrt{W} : N(\sqrt{m}, \sqrt{2m})$

④ $W : N(m, 2m), \quad \sqrt{W} : N(\sqrt{m}, 1)$

⑤ $W : N(m, 2m), \quad \sqrt{W} : N(\sqrt{m}, 1/2)$

問 10　$W(t),\ 0 \leq t,$ を標準ブラウン運動とする。$W(1) = 1.0$ が与えられたもとで $W(0.5)$ の条件つき期待値 (ア) と条件つき分散 (イ) はいくらか。正しい組合せを次の ① ～ ⑤ のうちから一つ選べ。 **19**

① ア: 0.0　イ: 0.5　　② ア: 0.5　イ: 0.5　　③ ア: 1.0　イ: 0.5
④ ア: 0.5　イ: 0.25　　⑤ ア: 0.0　イ: 0.25

問 11　映画館に関する経済産業省「平成 22 年特定サービス産業実態調査（確報）」のデータでは，島根県と徳島県のスクリーン数が秘匿されている。これは，県内の映画館事業所が 2 つしかないための措置である。

徳島県のスクリーン数を回帰式で予測することを考える。被説明変数を都道府県別スクリーン数 y，説明変数を都道府県別映画館従業者数 x として，回帰式 $y = \beta_0 + \beta_1 x$ を想定する。これを最小二乗法で推定し，徳島県の映画館従業者数 48 人を代入してスクリーン数を予測する。

〔1〕下に示す回帰式の推定結果に基づく徳島県のスクリーン数の予測値を，次の ① ～ ⑤ のうちから一つ選べ。 **20**

① 20.5　② 22.2　③ 23.6　④ 24.8　⑤ 26.1

計算結果

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-54.566  -9.848  -4.421   6.684  82.728

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.465565   3.862580   4.781 2.07e-05 ***
x            0.106381   0.004791  22.204  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.41 on 43 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.9198,    Adjusted R-squared:  0.9179
F-statistic:   493 on 1 and 43 DF,  p-value: < 2.2e-16
```

〔2〕下の散布図（図中の点線は，徳島県の映画館従業者数をあらわす）及び回帰診断図から判断して，上の回帰分析及び徳島県のスクリーン数の予測の問題点を述べよ。 **記述 7**

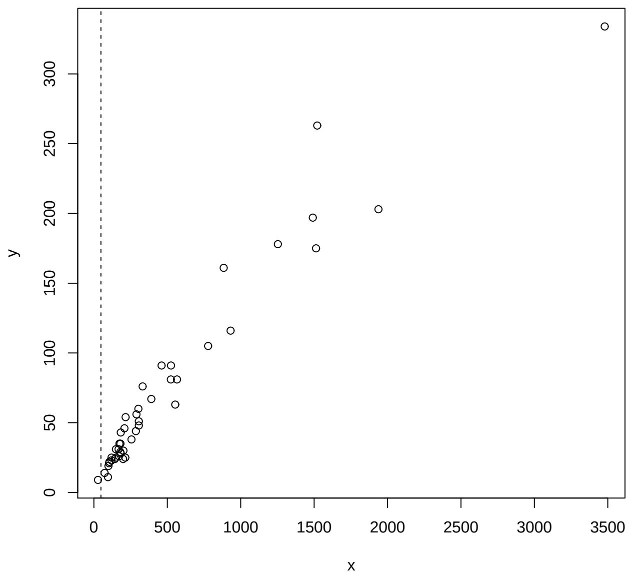


図 1: 都道府県別映画館従業者数 x と映画館スクリーン数 y の散布図
資料：経済産業省「平成 22 年特定サービス産業実態調査」（確報）映画館

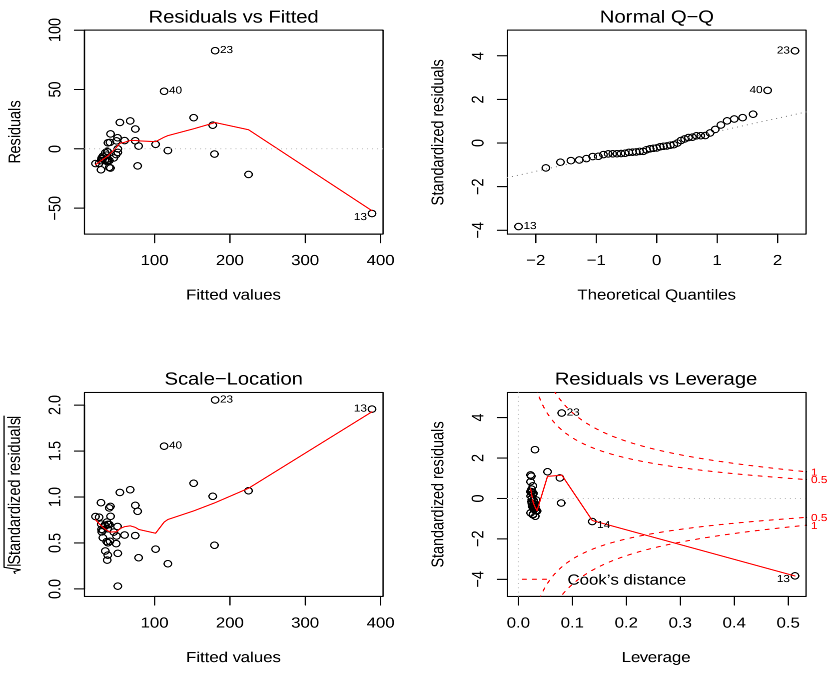
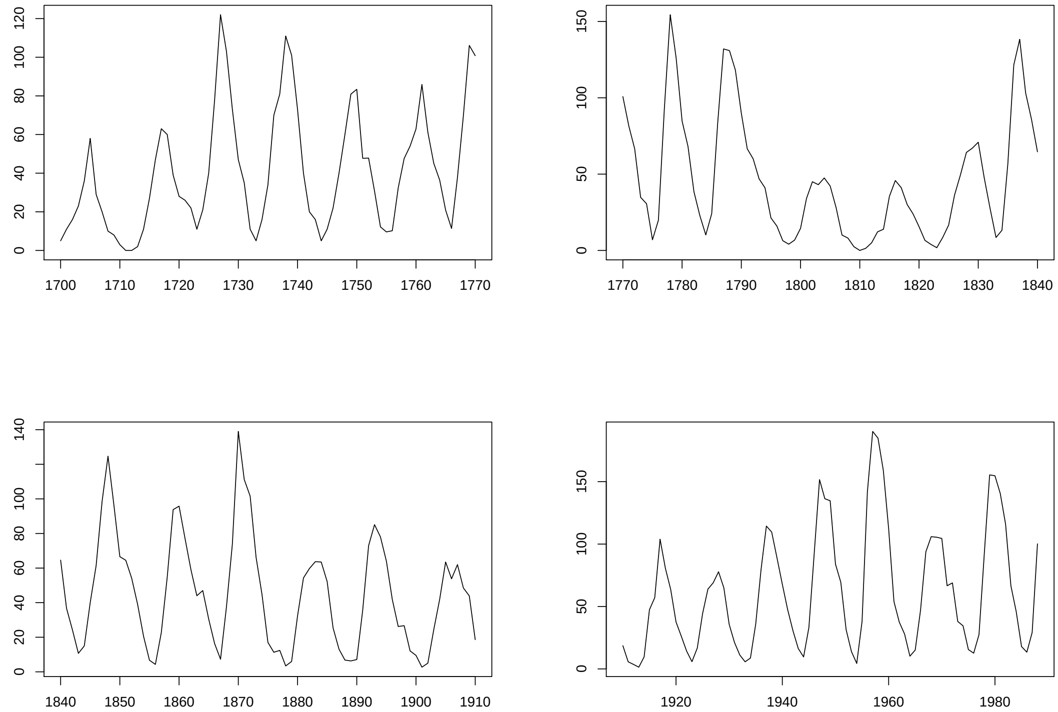


図 2: 回帰診断図

問 12　ある 20 人のクラスで，勉強時間 x とテストの成績 y との相関係数を計算したら 0.50 であった。このクラスを無作為に 10 人ずつに分割して，一方で x の平均を，もう一方で y の平均を計算する。（無作為分割を反復した際の）x の平均と y の平均の相関係数はいくらか，次の ① ～ ⑤ のうちから一つ選べ。 **21**

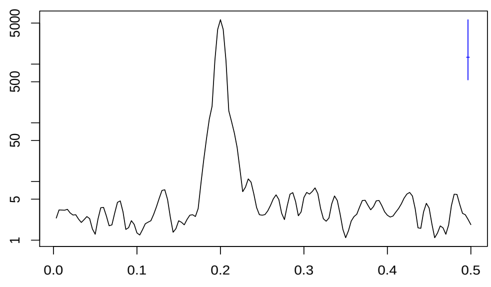
① -0.50　　② -0.05　　③ 0.0　　④ 0.05　　⑤ 0.50

問 13　次図の 4 つのグラフは太陽黒点の時系列の時系列プロットである (1700 年から 1988 年までをほぼ 70 年ごとに分割している)。

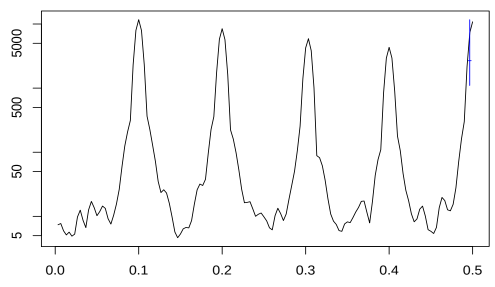


この時系列の平滑化したペリオドグラムとして適切なものを，次の ① ～ ⑤ のうちから一つ選べ。(各図中右上の縦棒は 95%信頼区間の長さを示す) $\boxed{\textbf{22}}$

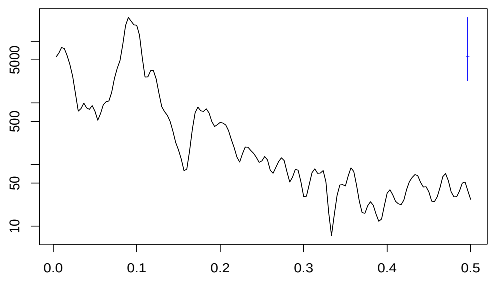
①



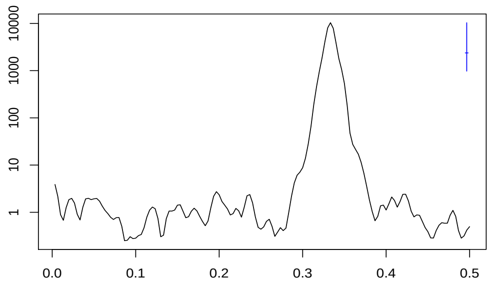
②



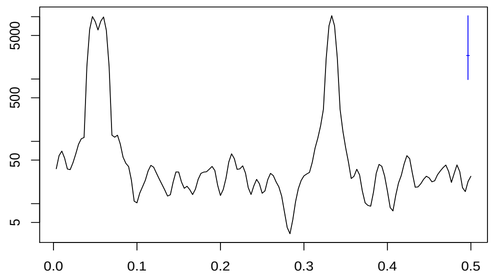
③



④



⑤



問 14　7 人が集まり麻雀 (4 人ゲーム) を 7 ゲーム行うことになった。ゲームが公平になるように，どの 2 人の参加者のペアについても，7 ゲーム中で対戦する回数が等しくなるような組合せを考える。これは，処理 (参加者) の数が 7，ブロックの大きさが 4，ブロックの数 (対戦の数) が 7 であるブロック計画のひとつで，釣合い型不完備ブロック計画とよばれる。

〔1〕このブロック計画において，各参加者の参加するゲーム数 (ア)，および参加者の各ペアの対戦数 (イ) の組合せとして正しいものを次の ① ～ ⑤ から一つ選べ。 **23**

① ア: 3　イ: 2　　② ア: 3　イ: 3　　③ ア: 4　イ: 1
④ ア: 4　イ: 2　　⑤ ア: 4　イ: 3

〔2〕この計画のブロックを部分的に以下で与えるとき，空欄のブロックとして適切な組合せを下の ① ～ ⑤ から一つ選べ。 **24**

1 回戦　$(1, 4, 6, 7)$
2 回戦　$(2, 4, 5, 6)$
3 回戦　$(1, 3, 4, 5)$
4 回戦　$(1, 2, 5, 7)$
5 回戦　$(3, 5, 6, 7)$
6 回戦　$\boxed{}$
7 回戦

① 6 回戦 $(1, 2, 3, 4)$, 7 回戦 $(2, 3, 6, 7)$

② 6 回戦 $(1, 2, 3, 4)$, 7 回戦 $(2, 5, 6, 7)$

③ 6 回戦 $(2, 3, 4, 6)$, 7 回戦 $(1, 2, 3, 5)$

④ 6 回戦 $(2, 3, 4, 6)$, 7 回戦 $(1, 2, 3, 7)$

⑤ 6 回戦 $(2, 3, 4, 7)$, 7 回戦 $(1, 2, 3, 6)$

問 15　A君は，ある路線の徒歩圏内にある一軒家の価格広告を利用しパス解析を試みた。物件の広告には価格（万円），新宿から最寄駅までの時間（分），最寄駅からの徒歩（分），土地面積（m^2），建物面積（m^2），築年数（年），部屋数（個）が示されている。

次の表は，A君がはじめにこれらの関係を仮定し，パス解析した結果（AMOS の出力）である。ここで,「<---」はA君が仮定した変数間の因果の関係であって，正しいかどうかは分からない。推定値，標準誤差はそれぞれ，変数間のパス係数に対するもので，検定統計量はパス係数が 0 であるという帰無仮説に対する t 検定の統計量である。確率は p-値を，*** は p-値が 0.001 未満であることを示す。

	推定値	標準誤差	検定統計量	確率
建物面積 <--- 部屋数	13.186	2.942	4.483	***
建物面積 <--- 土地面積	0.446	0.069	6.467	***
価格 <--- 築年数	-52.317	13.540	-3.864	***
価格 <--- 新宿から	-44.407	6.763	-6.566	***
価格 <--- 土地面積	10.395	2.769	3.754	***
価格 <--- 部屋数	-214.767	105.267	-2.040	.041
価格 <--- 建物面積	17.763	3.926	4.525	***
価格 <--- 徒歩	-33.918	23.793	-1.426	.154

〔1〕この結果について，A君は考え方の間違いに気づき修正した。その間違いは何であって，どのように修正したか。最も適切なものを次の ①～⑤ のうちから一つ選べ。 **25**

① 部屋数が建物面積の要因と考えるのはおかしいので，因果の関係を逆にした。

② 部屋数から価格への標準誤差が他との比較で大きすぎるので削除した。

③ p-値から徒歩が価格に関係しないことが分かったので削除した。

④ 部屋数は住居者の好みがあるので削除した。

⑤ 価格への直接のパスが多すぎるのでいくつかを削除した。

〔2〕〔1〕で示した間違いを正し，出力結果を整理した結果，以下のような関係がモデルの適合度の観点から最も良いものとして評価された。この結果をもとにパス図を作成せよ。 **記述 8**

	推定値	標準誤差	検定統計量	確率
建物面積 <--- 土地面積	.612	.076	8.014	***
価格 <--- 築年数	-55.921	14.043	-3.982	***
価格 <--- 新宿から	-43.829	7.014	-6.249	***
価格 <--- 土地面積	8.922	3.164	2.820	.005
価格 <--- 建物面積	15.725	3.676	4.278	***
部屋数 <--- 建物面積	.017	.003	6.560	***

問 16　$X_1, \ldots, X_n$ は互いに独立に正規分布 $N(\mu, \sigma^2)$ にしたがうとする。(μ, σ^2) の同時事前密度関数を

$$\pi(\mu, \sigma^2) \propto (\sigma^2)^{-2} \exp\left[-\frac{1}{2\sigma^2}(\mu^2+1)\right]$$

とすると，$x = (x_1, \ldots, x_n)'$ が観測されたとき (μ, σ^2) の同時事後密度関数は

$$\pi(\mu, \sigma^2|x) \propto (\sigma^2)^{-(n/2+2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\mu^2+1+\sum_{i=1}^{n}(x_i-\mu)^2\right\}\right]$$

となる。ただし $\bar{x} = \sum_{i=1}^{n} x_i/n$ は標本平均を，$\propto$ は両辺が比例的であることを，また $'$ はベクトルの転置を表す。

(μ, σ^2) の事後分布から，以下の初期化および**ステップ 1**–**ステップ 3** のようにギブス・サンプリングを用いて確率標本を発生させるとき，空欄 (A) に入る記述として適切なものを下の ①～⑤ のうちから一つ選べ。 **26**

初期化. (μ, σ^2) の初期値を $(\mu^{(0)}, \sigma^{2(0)})$ とし，$t=0$ とする。

ステップ 1. $\sigma^{2(t)}, x$ を所与として，$\mu^{(t+1)}$ を (A) から発生させる。

ステップ 2. $\mu^{(t+1)}, x$ を所与として，$\sigma^{2(t+1)}$ を以下の密度関数を持つ逆ガンマ分布から発生させる。

$$\pi(\sigma^2|\mu^{(t+1)}, x) \propto (\sigma^2)^{-(n/2+2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\left(\mu^{(t+1)}\right)^2+1+\sum_{i=1}^{n}\left(x_i-\mu^{(t+1)}\right)^2\right\}\right],$$

ステップ 3. t を $t+1$ としてステップ 1 に戻る。

① 平均 $\mu^{(t-1)}$, 分散 $\sigma^{2(t)}/n$ の正規分布

② 平均 $\bar{x}$, 分散 $\sigma^{2(t)}/n$ の正規分布

③ 平均 $\bar{x}$, 分散 $\sigma^{2(t)}/(n+1)$ の正規分布

④ 平均 $\frac{n}{n+1}\bar{x}$, 分散 $\sigma^{2(t)}/n$ の正規分布

⑤ 平均 $\frac{n}{n+1}\bar{x}$, 分散 $\sigma^{2(t)}/(n+1)$ の正規分布

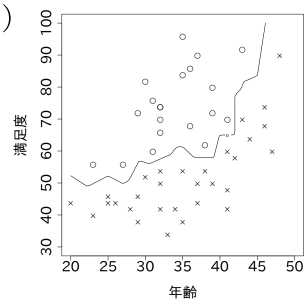
問 17 あるデパートで，商品購入者に対するアンケートを行い，年齢，満足度，今後も当デパートに再度来店したいかどうかについて回答してもらった。そして，年齢と満足度について基準化した変数を用い，今後も当デパートに来たいかどうかについて判別分析を行うこととした。

次の 5 つの図はアンケート結果のプロットと，線形判別，2 次判別，カーネル SVM，最近隣法，決定木のいずれかを用いた判別境界を示している（○印が「再度来店したい」，×印は「もう来店したくない」を表している）。ただし，線形判別と 2 次判別に関しては，事前確率はサンプルの比率を使うこととし，カーネル SVM ではガウシアンカーネルを用いている。

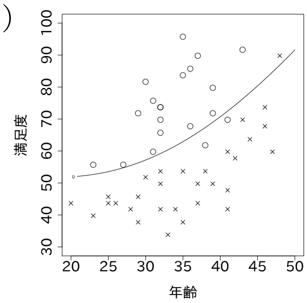
〔1〕以下の（ア）～（オ）の図の中で，2 次判別の判別境界を表しているものはどれか。次の ① ～ ⑤ のうちから適切なものを一つ選べ。 $\boxed{\textbf{27}}$

① （ア）　② （イ）　③ （ウ）　④ （エ）　⑤ （オ）

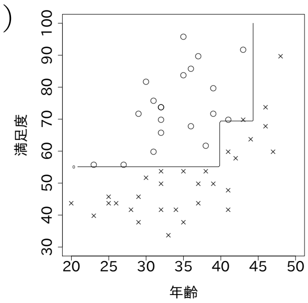
（ア）



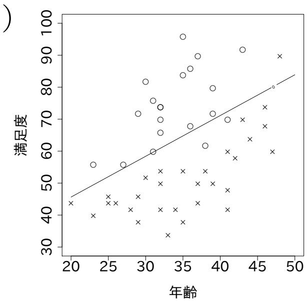
（イ）



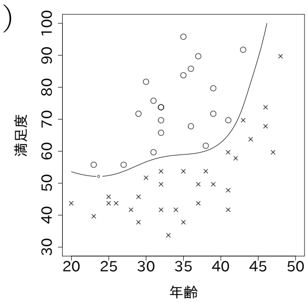
（ウ）



（エ）



（オ）



〔2〕5 種類の判別分析に関して述べた次の ① ～ ⑤ の意見のうちから，最も適切なものを一つ選べ。 **28**

① 線形判別は，分布が正規分布に従っていれば，2 群の分散共分散行列によらずに適用してよい。

② 2 次判別は，2 群の分散共分散行列が異なっていても適用できるが，分布は正規分布に従っていなければならない。

③ カーネル SVM は，指定すべきパラメータがいくつかあるが，分析結果にほとんど影響しないので，適当なパラメータを設定してよい。

④ 最近隣法は，データが特定の分布に従っているときにしか適用できない。

⑤ 決定木は，変数間に相関があるデータの分析に適している。

II. 論述問題

(3 題中 1 題選択)

問 1　あるチョコレート菓子の箱には，おまけとしてシールが 1 枚ずつ入っている。シールは全部で K 種類あり，菓子の箱に各シールが入っている確率は等しく $1/K$ ずつである。以下の各問に答えよ。

〔1〕$K=2$ とする。2 種類のシールの両方を手に入れるまでに購入する菓子の個数を X としたとき，確率 $P(X=x)$ を x の式として求めよ。また，X の期待値 $E[X]$ はいくらか。

〔2〕$K=3$ とする。3 種類のシールすべてを手に入れるまでに購入する菓子の個数を X としたとき，確率 $P(X=x)$ を x の式として求めよ。また，X の期待値 $E[X]$ はいくらか。

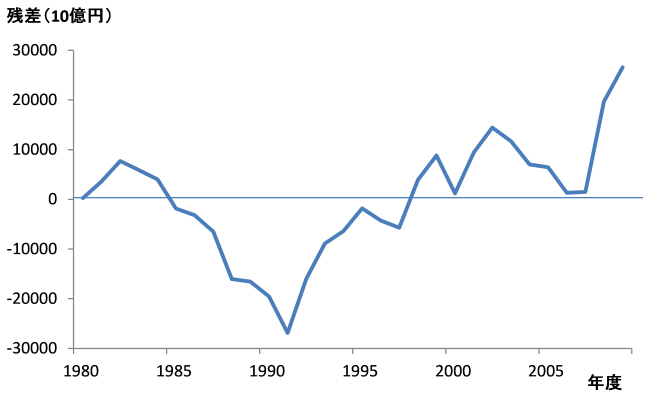
〔3〕$K=20$ のとき，すべての種類のシールを集めるためには，平均何個の菓子を買えばよいか。

問 2　1980 年度から 2009 年度までの実質民間最終消費支出を Y_t，実質国民可処分所得を X_t（いずれも単位 10 億円。$t = 1, 2, \ldots, 30$）として，被説明変数 Y を説明変数 X で単回帰したところ，次の推定結果が得られた。

$$\hat{Y}_t = \underset{(-3.66)}{-53800} + \underset{(19.82)}{0.8308}\, X_t, \qquad R^2 = 0.9335,\ \bar{R}^2 = 0.9311,\ s = 11896,\ AIC = 565.0$$

ただし，( ) 内は t 値，R^2 は決定係数，$\bar{R}^2$ は自由度修正済決定係数，s は誤差項の標準偏差の推定値，AIC は赤池の情報量基準である。この結果について，以下の〔1〕～〔3〕の問に答えよ。

〔1〕上の推定結果に基づいて各年度の残差 $e_t = Y_t - \hat{Y}_t$ を計算したところ，次の図のようになった。



この残差をもとに算出した 1 次の自己相関係数 $\hat{\rho}$（e_t と e_{t-1} の相関係数）とダービンワトソン統計量 $DW = \sum_{t=2}^{30}(e_t - e_{t-1})^2 / \sum_{t=1}^{30} e_t^2$ の値は $\hat{\rho} = 0.85$, $DW = 0.29$ であった。回帰モデルの通常の仮定に照らして，このような残差の検討の意味について述べよ。

〔2〕〔1〕の単回帰に人口 N（単位千人）を説明変数に加えて重回帰分析を行ったところ，次の推定結果が得られた。

$$\hat{Y}_t = \underset{(-12.55)}{-1119473} + \underset{(3.63)}{0.2013}\, X_t + \underset{(11.93)}{10.4302}\, N_t,$$
$$R^2 = 0.9894,\ \bar{R}^2 = 0.9886,\ s = 4838,\ AIC = 511.1$$

この推定結果について，それぞれの回帰係数の解釈について述べよ。また消費の所得弾力性を平均値で評価する方法について述べよ。

〔3〕実質可処分所得のみを説明変数とする単回帰分析と，これに人口を加えた重回帰分析のどちらのモデルがよりすぐれているかについて，回帰係数の検定およびモデル選択の観点から説明せよ。

問 3　ある薬剤の有害事象の発生率を薬剤の投与群と非投与群で比較する際，発生率の評価ではリスク比やオッズ比が用いられることが多い。投与群での事象の発生率を p_1 とし，非投与群での事象の発生率を p_0 とすると，リスク比 (Risk Ratio) は $RR = p_1/p_0$ で定義され，オッズ比 (Odds Ratio) は $OR = \{p_1/(1-p_1)\}/\{p_0/(1-p_0)\}$ で定義される。以下の各問に答えよ。

〔1〕リスク比 RR と非投与群での事象の発生率 p_0 を用いてオッズ比 OR を求める式を導け。

〔2〕観測データが

	発生	非発生	計
投与群	a	b	m
非投与群	c	d	n
計	s	t	N

と表わされるとき，オッズ比は通常 $OR^* = (ad)/(bc)$ と推定され，対数オッズ比 $\log OR^*$ の標準誤差は

$$SE[\log OR^*] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

で与えられる。これを参考に，リスク比 RR の推定値 RR^* と対数リスク比 $\log RR^*$ の標準誤差を求めよ。

〔3〕ある研究では，上問〔2〕の記号を用いると $m = 51$, $n = 50$ で $OR^* = 3.38$, $\hat{p}_0 = c/n = 0.58$ であったという（数値は小数第 3 位を四捨五入)。このとき，リスク比の推定値 RR^*，およびその 95%信頼区間を求めよ。

〔4〕上問〔3〕の結果を踏まえ，リスク比をオッズ比で近似する場合の注意点を述べよ。

選択問題及び部分記述問題略解

問 1　〔1〕：③

包除原理により $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ と求められる。

〔2〕：③

同様に

$$P(A{\cup}B{\cup}C) = P(A){+}P(B){+}P(C){-}P(A{\cap}B){-}P(A{\cap}C){-}P(B{\cap}C){+}P(A{\cap}B{\cap}C)$$

と求められる。

問 2　〔1〕：⑤

パラメータ λ のポアソン分布の確率関数は

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

であり，$k = 3$ を代入する。

〔2〕：④

パラメータ λ のポアソン分布の期待値と分散は $E(X) = \lambda$，$V(X) = \lambda$ であり，

$$V(X) = E(X^2) - \{E(X)\}^2$$

から求められる。

〔3〕：⑤

死亡者数は 0 人から 6 人以上までの 7 区分であるが，データの平均値 2.0 をもとにパラメータを推定しているため，検定統計量の自由度は $5(= 7 - 2)$ である。χ^2 値 4.498 を自由度 5 の χ^2 分布の上側 5 %点（11.07）を比較すると，$4.498 < 11.07$ なので，平均値 2.0 のポアソン分布にしたがっていることは棄却できない。

問 3　〔1〕：③

$E[X + Y] = 3.0$, $E[X - Y] = -1.0$, $V(X + Y) = 11.0$, $V(X - Y) = 3.0$, $Cov(X + Y, X - Y) = -1.0$ であることからわかる。

〔2〕：④

確率変数 (X, Y) が 2 変量正規分布

$$N\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2\end{pmatrix}\right)$$

に従う時，X が与えられた時の Y の条件つき分布は

$$N(\mu_Y + (\sigma_{XY}/\sigma_X^2)(X - \mu_X), \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2)$$

であることからわかる。

問 4 〔1〕: ②

標準正規分布の上側 2.5%点 $z_0 = 1.96$ 及び標準偏差 σ の推定値 $s = \sqrt{0.62(1-0.62)}$ を用いて，95%信頼区間は $(0,62 - z_0 \frac{s}{\sqrt{n}}, 0,62 + z_0 \frac{s}{\sqrt{n}})$ とする。

〔2〕: ①

信頼区間の幅を狭くするのは，n を大きくするか，z_0 を小さくするかである。信頼係数 $(1-\alpha)$ を大きくすると z_0 は大きくなる。

〔3〕: ③

平均 p，分散 $p(1-p)/n$ の正規分布で近似することにより，$P(X/n > 0.62) \approx 0.0968$ と求まる。

〔4〕: ④

$p(1-p)$ は $p = 0.5$ で最大になることを考慮し，

$$0.02 = 1.96\sqrt{\frac{0.5 \times 0.5}{n}}$$

を解いて $n \approx 2400$ を得る。

問 5 〔1〕 (ア) : 120, (エ) : 80

年齢層と商品タイプの選択が全く関係ない時，各年齢層と商品タイプに対する期待度数は

(年齢層の割合)×(商品タイプの割合)×(総人数)

で計算できる。

〔2〕(オ) : 4.17, (カ) : 3.84

分割表で，行と列が関係していないという仮説の検定統計量は

$$\sum_{\text{全セル}} \frac{\{(\text{セルの度数}) - (\text{セルの期待度数})\}^2}{(\text{セルの期待度数})}$$

で計算される。

問 6 : ③

第 1 列の列和から x_{11} は 10 以下である。第 2 行の行和から x_{21} は 8 以下であるため，x_{11} は 2 以上である。実際に，それぞれ $(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}) = (10, 2, 0, 0, 0, 8)$ と $(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}) = (2, 2, 8, 8, 0, 0)$ で実現する.

問 7 〔1〕: ②

箱ひげ図 b と c : どちらも四分位範囲の中に品種 A と品種 B の平均値があり，b では c に比べて中央値が小さく四分位範囲は大きい。正規分布を仮定しているので，b の方が c より平均値は小さく分散は大きいと判断できる。

〔2〕: 0.124

サンプルサイズ21の品種Aの分散と，サンプルサイズ16の品種Bの分散より，プールした分散は，

$$\frac{20 \times 0.145 + 15 \times 0.095}{21+16-2} = 0.1235\cdots \fallingdotseq 0.124$$

と計算できる。

〔3〕: -1.284

品種Aのサンプルサイズ n_A，品種Bのサンプルサイズ n_B であり，品種Aの平均値 $\bar{x}_A$，分散 s_A^2，品種Bの平均値 $\bar{x}_B$，分散 s_B^2，プールした分散 s^2 とすると，検定統計量である t 統計量は以下のように求められる。

$$t = \frac{\bar{x}_A - \bar{x}_B}{s\sqrt{\dfrac{1}{n_A}+\dfrac{1}{n_B}}} = \frac{2.78-2.93}{s\sqrt{\dfrac{1}{21}+\dfrac{1}{16}}}$$

〔4〕: ④

帰無仮説は「2つの品種のある部位の長さに差がない（ア）」，対立仮説は「2つの品種のある部位の長さに差がある」の両側検定を行う。棄却限界は，-2.030 および 2.030 である。t 統計量の値が -1.284 であり，この値が棄却限界の -2.030 より小さくないので，帰無仮説は棄却されない（イ）。したがって,「2つの品種のある部位の長さに差があるとはいえない（ウ）」と結論する。

問8 〔1〕: ①　〔2〕: ③　〔3〕: ①

〔1〕 ① 正しい。② 分散の信頼区間が0を含むことはない。③ 分散の信頼区間が1を含むことから有意でないことが判断できる。 ④ F検定の出力で示される信頼区間により p-値の大小がわかる。⑤ F-値は1から離れるほど分散の差異が大きいので，一般に F-値が小さい方が分散の差異が大きいとは言えない。

〔2〕 ①, ②, ④, ⑤ 正しい。③ このデータでは，Student の t 検定と Welch の t 検定の検定統計量が等しい。Student の t 検定の自由度が98であることから，Student の t 検定を用いたなら，有意になることがわかる。

〔3〕 多重比較について問う問題である。① 正しい。② 結果は順番に影響されない。③ 互いの分散が同じでも問題がある。④ 検出力は良くなっても悪くはならない。⑤ 検定に関する意見ではない。

問9：⑤

自由度 m のカイ二乗分布の平均は m 分散は $2m$ であり，m が大きい時は中心極限定理により正規分布で近似できる。デルタ法は X が分散 σ^2 の小さい正規分布 $N(\mu,\sigma^2)$ に従う時，$f(X)$ の分布を $N(f(\mu), f'(\mu)^2\sigma^2)$ で近似するものである。$\sqrt{W} = \sqrt{m}\sqrt{W/m}$ とおいて，$X = W/m$, $f(x) = \sqrt{x}$, $\mu = 1$, $\sigma^2 = 2/m$ としてデルタ法を応用することにより ⑤ が正しいことがわかる。

問10：④

$W(1)$ のしたがう分布は平均0，分散1の正規分布である。$(W(0.5), W(1))$ のしたがう同時分布は平均 $(0,0)$ の正規分布であり，$W(0.5)$ の分散は0.5，$W(1)$ の分散は1，

$W(0.5)$ と $W(1)$ の共分散は 0.5 である。したがって，$W(1)$ が与えられたもとでの $W(0.5)$ の条件つき分布は，平均 $W(1)/2$，分散 0.25 の正規分布である。

問 11 〔1〕：③　　予測値は $18.465565 + 0.106381 \times 48 = 23.57185$ である。

〔2〕 自由度調整済み決定係数や，従業者数に対する回帰係数が有意であることから，一見，導出された回帰式が有効と判断されるが，これは外れ値（都道府県番号 13）の影響が大きいためである。診断図からもこの外れ値の影響を考慮する要請がある。従業員数が多いほど誤差が大きくなる傾向があり，正規性も疑問である。そのため，誤差項の不均一分散および非正規性に対する処理が必要である。また，緩やかな曲線に沿っていることが見られるため，非線形の回帰式も考慮する。

問 12 ：①

無作為に分割された 10 人ずつのグループを，それぞれ A, B とする。また，A グループの x の平均を $\bar{X}_A$ とあらわし，他の記法も同様とする。20 人全体の平均を $\bar{x}, \bar{y}$ とあらわす。このとき，$\bar{Y}_B = 2\bar{y} - \bar{Y}_A$ となることに注意すれば,（無作為分割の反復に関して）$V(\bar{Y}_B) = V(\bar{Y}_A)$，$Cov(\bar{X}_A, \bar{Y}_B) = -Cov(\bar{X}_A, \bar{Y}_A)$ が得られる。したがって

$$\rho_{\bar{X}_A, \bar{Y}_B} = -Cov(\bar{X}_A, \bar{Y}_A)/\sqrt{V(\bar{X}_A)\,V(\bar{Y}_A)} = -\rho_{x,y}$$

となる。ただし，$\rho_{x,y}$ は 20 人全体で計算した x と y との相関係数である。

問 13 ：③

ペリオドグラムは，時系列の各周波数（1/周期）成分の大きさを表している。問題の太陽黒点の時系列では，10 年前後の明らかな周期が見られるため，ペリオドグラムは周波数 $1/10 = 0.1$ 前後でピークを取るべきであり，これを唯一満たしている ③ が最も適切であると考えられる。

問 14 〔1〕：④

このブロック計画では，[ブロックの数 7]×[ブロックのサイズ 4]÷ [参加者の数 7] = 4 なので，どの参加者も，4 回対戦を行う。会合数，すなわち任意の二人が対戦する回数，を λ とおく。ある参加者，例えば参加者 1 に注目すると，参加者 1 が含まれる 4 つの対戦（ブロック）では，参加者 1 以外の 3 人の対戦者が，全部で $4 \times 3 = 12$ 人必要である。この 12 人は，参加者 2∼ 参加者 7 の 6 人から選ばれるが，会合数の定義より，参加者 2∼ 参加者 7 のどの参加者も，参加者 1 との対戦数が等しいことから，$12 = 6 \times \lambda$ となる。従って，会合数は $\lambda = 2$ と定まる。

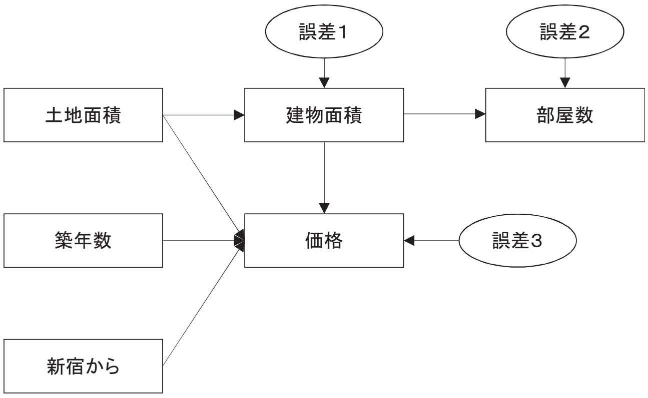
〔2〕：⑤

既に定まっている 1 回戦 ∼5 回戦を見ると，参加者 1 とまだ 2 回対戦していないのは，参加者 2，3，6 の 3 名であり，参加者 4，5，7 の 3 名は既に 2 回対戦している。従って，6 回戦と 7 回戦のいずれかは，$(1, 2, 3, 6)$ でなければならないことが分かる。そのような選択肢は ⑤ しかない。

問 15 〔1〕：①

① 正しい。
② 標準誤差は推定値に比べると大きくないため削除要因とはならず，誤り。（負の推定値により部屋数が多いほど価格が低くなってしまうことは削除要因となりうる。）
③ 徒歩から価格へ第 3 の変数を介して関係していることがあるため，誤り。
④ 部屋数が居住者の好みによることはあっても，それにより部屋数を削除する理由にはならず，誤り。
⑤ 価格への直接のパスが多いことが必ずしも間違いにはならないため，誤り。（直接のパスの推定値に多重共線性が見られる場合などは削除を検討することも考えられる。）

〔2〕以下の図のように，出力結果に従って変数間を矢印で結び，矢印が向けられた各変数に誤差の矢印を加えたものがパス図となる。



問 16 ：⑤

$\sigma^{2(t)}, x$ を所与とする μ の周辺事後密度関数は

$$\pi(\mu|\sigma^{2(t)}, x) \propto (\sigma^{2(t)})^{-(n/2+2)} \exp\left[-\frac{1}{2\sigma^{2(t)}}\left\{\mu^2 + 1 + \sum_{i=1}^{n}(x_i - \mu)^2\right\}\right]$$
$$\propto \exp\left[-\frac{n+1}{2\sigma^{2(t)}}\left(\mu - \frac{1}{n+1}\sum_{i=1}^{n} x_i\right)^2\right]$$

となることから，$\mu^{(t+1)}$ を発生させるべき (A) の分布は平均 $\frac{n}{n+1}\bar{x}$, 分散 $\sigma^{2(t)}/(n+1)$ の正規分布となる。

問 17 〔1〕：②

2 次判別の判別境界は一般に 2 次曲線となるため，境界線が明らかに 2 次曲線でない（ア）,（ウ）,（オ）は該当しない。（エ）のような直線の境界となる場合もありう

るが,（エ）よりも（イ）の方が当てはまりの良い判別境界であることから 2 次判別の結果は（イ）であると考えられる。

〔2〕: ②

① 線形判別が適切となるのは各群の分散共分散行列が等しい場合であるため，誤り。
② 正しい。
③ カーネル関数やペナルティ項のパラメータに依存して SVM の分析結果は変わるため，誤り。
④ 最近隣法は特定の分布形を仮定していないため，誤り。
⑤ 変数間の相関が高いと決定木の分解性能が悪くサイズが大きくなるため，誤り。

論述問題略解

問 1　〔1〕 $P(X=1)=0$ であり，$x \geq 2$ に対しては $P(X=x)=(0.5)^{x-1}$ である。期待値は $E(X)=3$ である。

〔2〕 求める確率は

$$P(X=x)=\left(\frac{2}{3}\right)^{x-1}-2\left(\frac{1}{3}\right)^{x-1}$$

となる。期待値は $E(X)=5.5$ である。

〔3〕

$$E(X)=1+20\left(\frac{1}{19}+\frac{1}{18}+\cdots+1\right)\approx 71.955$$

となる。

問 2　〔1〕回帰モデル $Y_t=\alpha+\beta X_t+\epsilon_t$ において,「誤差項は自己相関していない」という仮定が，標準的な仮定の 1 つとして設定される。実証的には，自己相関のうち 1 次の自己相関，すなわち ϵ_t と ϵ_{t-1} の相関が問題になる．1 次の自己相関を $\epsilon_t=\rho\epsilon_{t-1}+u_t$ とし，自己相関係数 ρ について，$\rho=0$ すなわち自己相関なしの仮定をダービンワトソン統計量 DW で検定する場合，誤差項に正の自己相関があれば DW は 0 に近い値を，負の自己相関があれば DW は 4 に近い値をとり，DW が 2 に近い値をとれば自己相関がないことになる。ここで，$DW=0.29$ であり，DW はかなり 0 に近く，誤差項の自己相関なしの仮説 $\rho=0$ は棄却される。また，残差のグラフをみると，残差が負の値をとる期間が連続したり（1986–1998 年度），正の値をとる期間が連続したり（1999 年度以降），1 期前の残差と当期の残差に相関がある場合の典型的なグラフとなっている。

〔2〕実質国民可処分所得の係数は 0.2013 であり，これは人口を一定とした場合，可処分所得が 10 億円増加すると，実質民間最終消費支出が約 2 億円増加することを意味する。人口の係数は 10.4302 であり，これは実質国民可処分所得を一定とすると，人口が 1000 人増加すると，消費が 104.302 億円増加することを表す。消費の所得弾力性 η は，$\eta=\frac{dY/Y}{dX/X}=\beta\div\frac{Y}{X}$ と定義できる（β は所得の係数)。したがって，この弾力性を平均値で評価するには，$\beta\div\frac{\bar{Y}}{\bar{X}}$ とすればよい。したがって，β の推定値である 0.2013 を，実質民間最終消費支出と実質国民可処分所得のそれぞれの平均値から算出した可処分所得に対する消費の割合（平均消費性向）で除すれば，平均値で評価した消費の所得弾力性 η を算出できる。

〔3〕説明変数を追加すると R^2 の値は必ず大きくなる（少なくとも減少しない）ので，R^2 で両者のあてはまりの程度を比較することはできない。そこで自由度修正済決定係数 $\bar{R}^2$ でみると，$\bar{R}^2$ は 0.9311 から 0.9886 に上昇しており，これは人口 N を加えたことにより説明力が上昇したことを意味している。また，AIC をみると 565.0 から 511.1 に低下している。AIC は説明変数の個数と誤差項の標準偏差を総合的に評価したモデル選択の指標であり，AIC が小さい方が

望ましいモデルと判断できる。よって，$\bar{R}^2$ と AIC からみると，人口 N を加えた重回帰の方が望ましい。

さらに，追加した人口 N の係数の有意性を検定すると，N の係数の t 値が 11.9279 と十分大きいので，N の係数が 0 であるという仮説は棄却される（データの個数が 30 なので，正規分布による検定を利用しても問題ないため，5 %点は 1.645 である）。また，実質国民可処分所得 X の係数の t 値も 1.645 を上回っており，単回帰の場合と同様に有意である。

以上から，人口 N を加えた重回帰分析の方がすぐれていると言える。

問 3　〔1〕

$$OR = \frac{p_1}{p_0} \times \frac{1-p_0}{1-p_1} = RR \times \frac{1-p_0}{1-p_0 RR}$$

〔2〕リスク比は $RR^* = \frac{a/m}{c/n}$ で推定され，標準誤差は

$$SE[\log RR^*] = \sqrt{\frac{1}{a} - \frac{1}{m} + \frac{1}{c} - \frac{1}{n}}$$

となる。

〔3〕 $\hat{p}_0$ と OR^* より $\hat{p}_1 = 0.82$ なので $RR^* = 1.42$ であり，95%信頼区間は $(1.09, 1.86)$ となる。

〔4〕 事象の発生率がある程度大きい場合は，オッズ比はリスク比をかなりの程度過大評価する。

統計検定 準1級 部分記述用　解答用紙（例）

問5

記述1	記述2	記述3	記述4

問7

記述5	記述6

問11

記述7

問15

記述8