

平成30(2018)年度
東京大学大学院学際情報学府学際情報学専攻
(生物統計情報学コース)
入学試験問題
専門科目

(平成30年1月16日14:00~16:00)

試験開始の合図があるまで問題冊子を開いてはいけません。開始の合図があるまで、下記の注意事項をよく読んでください。

1. 本冊子は、生物統計情報学コースの受験者のためのものである。
2. 本冊子の本文は8ページである。落丁、乱丁、印刷不鮮明の箇所などがあった場合には申し出ること。
3. 本冊子には、第1問から第3問までの計3問の問題が収録されている。第1問は択一式問題であり、全員が解答すること。第2問及び第3問は記述式問題であり、この2問の中から1問を選択して解答すること。
4. 本冊子の問題は、日本語文で記述されている。
5. 解答用紙は2枚ある。解答した問題ごとに解答用紙1枚を使用すること。このほかにメモ用紙が1枚ある。なお、解答用紙のみが採点の対象となる。
6. 解答用紙の上方の欄に、解答した問題の番号及び受験番号を必ず記入すること。問題番号及び受験番号を記入していない答案は無効である。
7. 解答には必ず黒色鉛筆(または黒色シャープペンシル)を使用すること。
8. 解答は日本語によるものとする。
9. 試験開始後は、中途退場を認めない。
10. 本冊子、解答用紙、メモ用紙、及び分布表が印刷された用紙は持ち帰ってはならない。
11. 次の欄に受験番号と氏名を記入せよ。

受験番号	
氏名	

生物統計情報学 第1問 (必須問題)

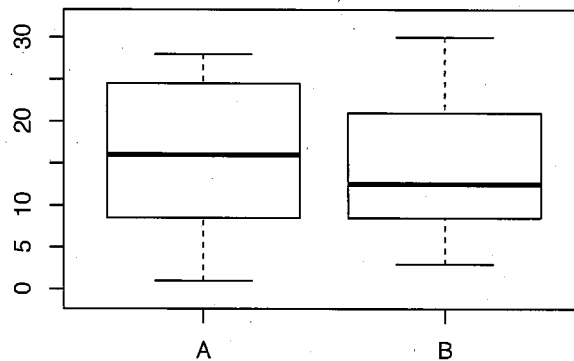
以下の問い(1-1)～(1-20)に答えよ。解答用紙には問いの番号と解答のみを、問いの番号の順序に従って記載せよ。なお、別途配布した分布表は適宜参照してよい。

(1-1) 次の幹葉図から読み取れることとして正しいものを、以下のア～オのうちから一つ選べ。

十の位	一の位
4	0
5	2 5
6	0 4 4 7
7	1 3 3 5 8
8	0 0 2 3 3 4 7
9	0 1 1 2 2 5

- ア. 最小値は45である。
- イ. 中央値は78である。
- ウ. 最大値は90である。
- エ. 範囲は95である。
- オ. 平均値は中央値よりも小さい。

(1-2) 次の二つの箱ひげ図A, Bを比較した記述として正しいものを、以下のア～オのうちから一つ選べ。



- ア. 最大値はAの方が大きい。
- イ. 中央値は等しい。
- ウ. 第1四分位数は等しい。
- エ. 四分位範囲は等しい。
- オ. 最小値はBの方が小さい。

(1-3) 次の(a)~(c)は、相関関係に関する記述である。適切な記述をすべて選んだ組合せを、以下のア~オのうちから一つ選べ。

- (a) 一方の変数の値が大きいとき、他方の変数の値も大きい傾向にあるとき、正の相関関係があるという。
- (b) 相関係数の値を確認することで、2変数の関係を線形・非線形問わず把握することができる。
- (c) 強い正の相関関係が観察された場合、一方の変数に介入することで他方の変数を制御することが必ずできる。

ア. (a) イ. (b) ウ. (c) エ. (a), (b) オ. (a), (c)

(1-4) 次の(a)~(c)は、母集団からのデータ抽出手法に関する記述である。適切な記述をすべて選んだ組合せを、以下のア~オのうちから一つ選べ。

- (a) 性別や職業などの基準によって母集団が層別されているとき、層ごとにランダムに抽出する手法をクラスター抽出法という。
- (b) 系統抽出法では最初に母集団に通し番号を付与するが、その際には番号の付け方をランダムにする必要がある。
- (c) 多段抽出法では、抽出単位を複数に分けて、段階的に抽出を行う。

ア. (a) イ. (b) ウ. (c) エ. (a), (c) オ. (b), (c)

(1-5) 1から6までの六つの目が等確率で出るサイコロを二つ投げたとき、二つのサイコロの目が異なる確率はいくらか。次のア~オのうちから正しいもの一つ選べ。

ア. $\frac{1}{36}$ イ. $\frac{1}{6}$ ウ. $\frac{1}{3}$ エ. $\frac{2}{3}$ オ. $\frac{5}{6}$

(1-6) 男性6人と女性2人の被験者計8人を4人ずつの二つの群にランダムにわけたとき、一方の群がすべて男性である確率はいくらか。最も近い値を次のア~オのうちから一つ選べ。

ア. 0.2 イ. 0.3 ウ. 0.4 エ. 0.5 オ. 0.6

(1-7) あるゲームに参加した 100 人の平均スコアは 60 点であった。さらに、スコアが 70 点の人の標準化得点は 2.5 であったことが知られている。スコアが 40 点であった人の標準化得点はいくらか。次のア～オのうちから正しいものを一つ選べ。

ア. -10 イ. -5 ウ. 0 エ. 5 オ. 10

(1-8) 昨年にある病院で生まれた新生児の身長について、平均は 48 cm、標準偏差は 3 cm であった。このデータに対する変動係数の値を CV とする。新生児それぞれの身長が 4 cm ずつ高かった場合、変動係数の値はもとの CV の何倍になるか。次のア～オのうちから正しいものを一つ選べ。

ア. $\frac{3}{4}$ イ. $\frac{12}{13}$ ウ. 1 エ. $\frac{13}{12}$ オ. $\frac{4}{3}$

(1-9) あるアンケートの結果、次のクロス集計表が得られた。この表から計算したオッズ比が 1 であるとき、 n の値として正しいものを、以下のア～オのうちから一つ選べ。

40	50
20	n

ア. 10 イ. 25 ウ. 30 エ. 55 オ. 70

(1-10) 二つの事象 A, B について次に示す (a)～(c) の式を考える。任意の事象 A, B に対して成立する式をすべて選んだ組合せを、以下のア～オのうちから一つ選べ。

(a) $P(A \cap B) = P(A)P(B|A)$

(b) $P(A \cup B) = P(A) + P(B)$

(c) $P(A \cap B) = P(A)P(B)$

ア. (a) イ. (b) ウ. (c) エ. (a), (b) オ. (b), (c)

- (1-11) 1から6までの六つの目が等確率で出る大小二つのサイコロがある。それぞれのサイコロを1回ずつ投げた結果について、次のA, B, Cという事象を考える。互いに独立な事象の対をすべて選んだ組合せを、以下のア～オのうちから一つ選べ。

A: 大きいサイコロの目が4以上である。
B: 小さいサイコロの目が4以下である。
C: 二つのサイコロの目の積が偶数である。

- ア. AとB
イ. AとC
ウ. BとC
エ. AとB, AとC
オ. AとB, BとC

- (1-12) ある疾患の有病確率が5%であったとする。ある血液検査では、この疾患を有している人の95%、有していない人の5%が陽性となる。ランダムに1名の対象者を選び、この血液検査を行ったところ、陽性となった。この対象者が疾患を有している確率はいくらか。最も近い値を次のア～オのうちから一つ選べ。

- ア. 0.05 イ. 0.10 ウ. 0.45 エ. 0.50 オ. 0.95

- (1-13) 確率変数 X の確率密度関数が次のように与えられている：

$$f(x) = \begin{cases} e^{-x+1} & (x > 1) \\ 0 & (x \leq 1) \end{cases}$$

確率変数 X の期待値として正しいものを、次のア～オのうちから一つ選べ。

- ア. 1 イ. 2 ウ. e エ. $2e$ オ. $\frac{1}{e}$

- (1-14) 確率変数 X_1, X_2 がそれぞれ平均 λ_1, λ_2 のポアソン分布に独立に従うとする。確率変数 $X_1 + X_2$ の従う確率分布の分散として正しいものを、次のア～オのうちから一つ選べ。

- ア. $\lambda_1 + \lambda_2$
イ. $(\lambda_1 + \lambda_2)^2$
ウ. $\frac{1}{\lambda_1 + \lambda_2}$
エ. $\frac{1}{(\lambda_1 + \lambda_2)^2}$
オ. $\frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$

(1-15) ある地域の中学生在が一ヶ月に読む本の冊数が、平均7冊、標準偏差4冊の確率分布に従うとする。この地域から無作為に抽出された中学生100人について、一ヶ月に読む本の合計冊数が700以上720以下である確率はいくらか。最も近い値を次のア～オのうちから一つ選べ。

ア. 0.01 イ. 0.10 ウ. 0.20 エ. 0.30 オ. 0.45

(1-16) 分散既知の正規分布 $N(\mu, 1)$ に対して、平均 μ の区間推定を考える。サンプルサイズを2倍に増やしたとき、信頼係数95%の信頼区間はどのように変化すると考えられるか。次のア～オのうちから適切なものを一つ選べ。

- ア. 変化しない。
- イ. 信頼区間幅がおよそ1/2倍になる。
- ウ. 信頼区間幅がおよそ $1/\sqrt{2}$ 倍になる。
- エ. 信頼区間幅がおよそ $\sqrt{2}$ 倍になる。
- オ. 信頼区間幅がおよそ2倍になる。

(1-17) 確率変数 X_1, \dots, X_n が平均 μ 、分散 σ^2 の正規分布 $N(\mu, \sigma^2)$ に独立に従うとする。このとき、 $W = \sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ が従う確率分布として正しいものを、次のア～オのうちから一つ選べ。ただし、 $\bar{X} = \sum_{i=1}^n X_i / n$ である。

- ア. 標準正規分布
- イ. 自由度 $n-1$ の t 分布
- ウ. 自由度 n の t 分布
- エ. 自由度 $n-1$ の χ^2 分布
- オ. 自由度 n の χ^2 分布

(1-18) 次の(a)～(c)は仮説検定に関する記述である。適切な記述をすべて選んだ組合せを、以下のア～オのうちから一つ選べ。

- (a) 第1種の過誤とは、帰無仮説が正しいにもかかわらず棄却してしまうことを意味する。
- (b) 第2種の過誤確率は、第1種の過誤確率に依存しない。
- (c) 帰無仮説が十分小さな有意水準のもとで棄却されなかった場合、帰無仮説が正しいことに対する有力な証拠が得られたことを意味する。

ア. (a) イ. (b) ウ. (c) エ. (a), (c) オ. (b), (c)

(1-19) ある地域の小学生 10,000 人の体重が、平均 40 kg、標準偏差 8 kg の正規分布に近似的に従うとする。この地域から 10 名の生徒を無作為抽出するとき、抽出された生徒の平均体重が 35 kg 以下である確率はいくらか。最も近い値を次のア～オのうちから一つ選べ。

ア. 0.01 イ. 0.02 ウ. 0.10 エ. 0.25 オ. 0.35

(1-20) n 個のデータ $(x_1, y_1), \dots, (x_n, y_n)$ から得られた回帰直線 $y = mx + 3$ が、点 $(3, 6)$ を通るとする。各変数の標本平均を $\bar{x} = \sum_{i=1}^n x_i/n$ および $\bar{y} = \sum_{i=1}^n y_i/n$ と表す。 \bar{x} を用いて \bar{y} を表した式として正しいものを、次のア～オのうちから一つ選べ。

- ア. \bar{x}
- イ. $\bar{x} - 2$
- ウ. $\bar{x} - 3$
- エ. $\bar{x} + 3$
- オ. $3\bar{x} + 3$

生物統計情報学 第2問 (選択問題)

確率変数 X と Y の同時確率密度関数が次のように与えられている：

$$f_{X,Y}(x,y) = \begin{cases} cxy & (0 < x < y < 2) \\ 0 & (\text{otherwise}) \end{cases}$$

ただし、 c は定数とする。このとき、以下の問いに答えよ。

(2-1) 定数 c を求めよ。

(2-2) 条件付き確率 $P(0 < X < 1/4 \mid Y = 1/2)$ を求めよ。

(2-3) 確率変数 X と Y は独立であるか。理由とあわせて述べよ。

(2-4) 確率変数 X, Y の相関係数 r_{XY} を求めよ。

(2-5) 実数 a, b, c, d を用いて、 $U = aX + b, V = cY + d$ という変数変換を考える。このとき、確率変数 U, V の相関係数 r_{UV} と (2-4) で求めた相関係数 r_{XY} の正負はどのような場合に逆であるか。実数 a, b, c, d に関する条件として求めよ。

生物統計情報学 第3問 (選択問題)

大腸がん発症予防における低用量アスピリンの効果を調べるためのランダム化プラセボ対照比較試験を考える。1,000名の対象者を低用量アスピリン群とプラセボ群へランダムに半数ずつ割り付け、全対象者を6年間追跡した。その結果、低用量アスピリン群では14名、プラセボ群では26名の大腸がん発症者が観察された。なお、プラセボとは、低用量アスピリンと外見上見分けのつかない、有効成分を含有しない偽薬のことである。このとき、以下の問いに答えよ。なお、別途配布した分布表は適宜参照してよい。

- (3-1) 低用量アスピリンの効果の有無について、統計的仮説検定により検討したい。この検定における帰無仮説および対立仮説を、適当な記号を定義したうえで明示せよ。
- (3-2) (3-1) で定義した帰無仮説に対する適切な検定手法はどのような手法か。その検定手法を選択した理由とともに、適当な記号を定義したうえでその検定統計量の式を書け。また、上記データに対するその検定統計量の値を求めるとともに、統計学的有意性を判断せよ。ただし、統計的仮説検定における有意水準は5%とする。
- (3-3) 低用量アスピリンの大腸がん発症予防効果に関して、(3-2)での解析結果をどのように解釈すべきか。自分の意見を120字以上200字以内で述べよ。