平成３１（２０１９）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８月２０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１０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) リンパ節への転移は悪性腫瘍の重症度に影響する．所属リンパ節が評価不能であるものを除くと，転移の程度に応じた分類は次のようになる．リンパ節転移に関する変数をこの分類から定義するとき，その尺度としてより適切なものを，以下のア〜オのうちから一つ選べ．

0：所属リンパ節への転移なし
1：所属リンパ節に1個の転移あり
2：所属リンパ節に2個以上の転移あり

ア．名義尺度

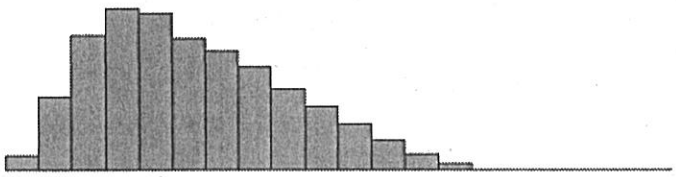
イ．順序尺度

ウ．間隔尺度

エ．比例尺度

オ．連続尺度

(1-2) 次の図は，ある糖尿病患者集団における HbA1c のヒストグラムである．図に対する平均値，中央値，最頻値を小さい順に左から並べたものとして適切なものを，以下のア〜オのうちから一つ選べ．



ア．平均値，中央値，最頻値

イ．最頻値，中央値，平均値

ウ．中央値，平均値，最頻値

エ．最頻値，平均値，中央値

オ．平均値，最頻値，中央値

(1-3) Quality of life (QOL) に関するある質問票の得点 X が，平均 50，分散 100 の正規分布に従っている．このとき，X が 60 点以下である人は何 % いるか．最も近い値を次のア～オのうちから一つ選べ．

ア．10%　　イ．16%　　ウ．32%　　エ．68%　　オ．84%

(1-4) 次の表は，二つの変数 X, Y に関する 6 人のデータをまとめたものである．この表から計算される X と Y の相関係数として正しいものを，以下のア～オのうちから一つ選べ．

	1	2	3	4	5	6
X	−5	−3	−1	1	3	5
Y	13	−2	−11	−11	−2	13

ア．−1　　イ．−0.5　　ウ．0　　エ．0.5　　オ．1

(1-5) 治療 A と治療 B のいずれかを，来院した 6 人の患者に順に割り付ける．治療 A と治療 B の患者が 3 人ずつになるようにランダムに治療を割り付けるとき，最初の 3 人に治療 A，残る 3 人に治療 B が割り付けられる確率はいくらか．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．0.05　　エ．0.10　　オ．0.17

(1-6) 1から6までの六つの目が出るサイコロを二つ投げるとき，出る目の組(X, Y)の同時確率が次の表で与えられたとする．二つのサイコロを投げたとき，$X = 2$であることがわかっているとすると，$Y = 1$である確率はいくらか．最も近い値を以下のア～オのうちから一つ選べ．

	Y=1	Y=2	Y=3	Y=4	Y=5	Y=6
X=1	0.10	0.02	0.03	0.02	0.01	0.01
X=2	0.02	0.08	0.03	0.03	0.01	0.01
X=3	0.01	0.02	0.05	0.03	0.01	0.01
X=4	0.01	0.01	0.03	0.05	0.02	0.01
X=5	0.01	0.01	0.03	0.03	0.08	0.02
X=6	0.01	0.01	0.02	0.03	0.02	0.10

ア．0.02　イ．0.03　ウ．0.11　エ．0.13　オ．0.17

(1-7) 10本のくじの中に当たりが3本ある．くじは順番に引き，一度引いたくじは戻さない．まだ誰もくじを引いていないとき，2番目の人が当たりを引く確率はいくらか．次のア～オのうちから正しいものを一つ選べ．

ア．$\frac{2}{9}$　イ．$\frac{3}{9}$　ウ．$\frac{2}{10}$　エ．$\frac{3}{10}$　オ．$\frac{5}{18}$

(1-8) ある確率pに対する対数オッズ$\log(p/(1-p))$のpに関する一階微分として正しいものを，次のア～オのうちから一つ選べ．

ア．$\frac{1}{p}$

イ．$\frac{1}{1-p}$

ウ．$\frac{p}{1-p}$

エ．$\frac{1}{p} + \frac{1}{1-p}$

オ．$\frac{1}{p} - \frac{1}{1-p}$

(1-9) 正規分布 $N(\mu, \sigma^2)$ のモーメント母関数 $M(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$ の導関数として正しいものを，次のア～オのうちから一つ選べ．

ア．$\mu t + \dfrac{\sigma^2 t^2}{2}$

イ．$\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$

ウ．$\mu t + \dfrac{\sigma^2 t^2}{2} + \exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$

エ．$(\mu + \sigma^2 t)\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$

オ．$\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$

(1-10) 確率変数 X の確率密度関数が

$$f(x) = \begin{cases} cx^2(1-x)^2 & (0 \le x \le 1) \\ 0 & (\text{otherwise}) \end{cases}$$

であるとき，c の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．6　　エ．12　　オ．30

(1-11) あるガンマ分布の規格化定数を求めるために $\int_0^\infty t^{-1/2}\exp(-t)dt$ を計算した結果として正しいものを，次のア～オのうちから一つ選べ．

ア．$\dfrac{1}{2}$　　イ．1　　ウ．$\sqrt{\pi}$　　エ．$\sqrt{e}$　　オ．e

(1-12) 互いに独立な二つの確率変数 X, Y が，$X \sim N(3, 9), Y \sim N(2, 4)$ と正規分布に従うとする．このとき，$X^2 + Y^2$ 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．5　　イ．10　　ウ．13　　エ．18　　オ．26

(1-13) 0から1までの連続値をとるVisual Analog Scale (VAS)により，100人の患者の痛みの程度を調べる．VASに関する確率変数$X_1, \ldots, X_{100}$が，区間$[0, 1]$上の一様分布に独立に従うとき，$\sum_{i=1}^{100} X_i/100$が0.55以上となる確率はいくらか．最も近い値を次のア〜オのうちから一つ選べ．ただし，$\sqrt{2} \approx 1.414$，$\sqrt{3} \approx 1.732$，$\sqrt{5} \approx 2.236$である．

ア．1%　　イ．2%　　ウ．4%　　エ．8%　　オ．16%

(1-14) 死亡するまでの時間Tが，確率密度関数$f(t) = 0.3\exp(-0.3t)$ $(t \geq 0)$で表される指数分布に従うとする．時点t_1まで生存することがわかっているとき，時点t_2 $(t_1 < t_2)$まで生存する確率はいくらか．次のア〜オのうちから正しいものを一つ選べ．

ア．$0.3\exp(-0.3t_1)$

イ．$0.3\exp(-0.3t_2)$

ウ．$\exp(-0.3t_1)$

エ．$\exp(-0.3t_2)$

オ．$\exp(-0.3(t_2 - t_1))$

(1-15) n人の患者について，治療の成功と失敗に関する確率変数$X_1, \ldots, X_n$が成功確率p $(0 < p < 1)$のベルヌーイ分布に独立に従うとする．nが十分大きいとき，近似的に標準正規分布に従う統計量として適切なものを，次のア〜オのうちから一つ選べ．ただし，$\overline{X}_n = \sum_{i=1}^{n} X_i/n$である．

ア．$\dfrac{\overline{X}_n - p}{p(1-p)}$

イ．$\dfrac{\overline{X}_n - p}{\sqrt{p(1-p)}}$

ウ．$\dfrac{\overline{X}_n - p}{n\sqrt{p(1-p)}}$

エ．$\dfrac{\overline{X}_n - p}{\sqrt{np(1-p)}}$

オ．$\dfrac{\overline{X}_n - p}{\sqrt{p(1-p)/n}}$

(1-16) 帰無仮説を $H_0 : \mu = 0.6$，対立仮説を $H_1 : \mu = 0.7$ とした仮説検定を考える．有意水準は 5% とし，採用した検定手法の検出力は 80% であるとする．この仮説検定に関する記述として適切なものを，次のア～オのうちから一つ選べ．

ア．第 1 種の過誤確率は 0.5 である．

イ．第 1 種の過誤確率は 0.8 である．

ウ．第 2 種の過誤確率は 0.5 である．

エ．第 2 種の過誤確率は 0.8 である．

オ．対立仮説 H_1 が正しいとき，帰無仮説 H_0 が棄却されない確率は 0.2 である．

(1-17) ある患者集団における治療 C の効果を知るために，ランダムに選ばれた 100 人の患者に治療 C を行った結果，100 人中 30 人が治癒した．このデータから計算される，治療 C の治癒確率に対する近似的な 95% 信頼区間として適切なものを，次のア～オのうちから一つ選べ．ただし，$\sqrt{2} \approx 1.414$，$\sqrt{3} \approx 1.732$，$\sqrt{5} \approx 2.236$，$\sqrt{7} \approx 2.646$ である．

ア．$(0.21, 0.39)$

イ．$(0.25, 0.35)$

ウ．$(0.29, 0.31)$

エ．$(0.61, 0.79)$

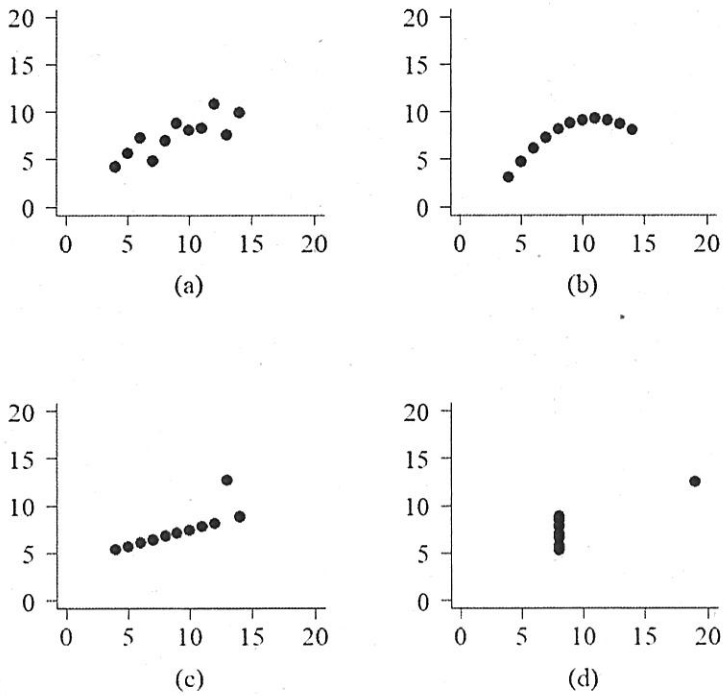
オ．$(0.65, 0.75)$

(1-18) ある会社の男女に普段の飲酒に関するアンケートを行った結果，次の表が得られた．この表から女性を基準として飲酒ありの確率に関するオッズ比を計算した結果として最も近い値を，次のア～オのうちから一つ選べ．

	飲酒あり	飲酒なし	合計
男性	132	88	220
女性	108	192	300
合計	240	280	520

ア．0　　イ．0.26　　ウ．1　　エ．1.78　　オ．2.67

(1-19) 次の四つの散布図において，ピアソンの相関係数 ρ は 0.82 であり，帰無仮説 $H_0 : \rho = 0$ に対する検定の p 値は 0.002 であった．これらの散布図に関する記述として適切なものを，以下のア～オのうちから一つ選べ．



資料：Anscombe FJ, Am Stat 1973

ア．$\rho = 0.82$ なので，すべての散布図で x が増加すれば y も増加する．

イ．p 値が 0.002 なので，すべての散布図で x が増加すれば y も増加する．

ウ．相関係数が x と y の直線の関係を反映しているのは (a) である．

エ．相関係数に関わらず，x と y が直線の関係にあると考えられる散布図は (a) と (b) のみである．

オ．相関係数に関わらず，x と y が直線の関係にあると考えられる散布図は (c) と (d) のみである．

(1-20) 1,000 人のデータから，テレビ視聴時間 (TV) を説明変数として，収縮期血圧 (SBP) を説明する線形回帰モデルを推定した結果，次の出力が得られた．この出力結果から考えられる記述として適切なものを，以下のア〜オのうちから一つ選べ．

```
Call:
lm(formula = SBP ~ TV)

Residuals:
     Min       1Q   Median       3Q      Max
-29.6304  -5.4643  -0.1049   5.3252  29.8892

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 124.7933     0.2297  543.39   <2e-16 ***
TV            0.3149     0.1075    2.93   0.0034 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.991 on 9998 degrees of freedom
Multiple R-squared:  0.0008578, Adjusted R-squared:  0.0007579
F-statistic: 8.584 on 1 and 9998 DF,  p-value: 0.003399
```

ア．Intercept の Pr(> $|t|$) が 2e-16 なので，SBP の変動は Intercept により十分説明される．

イ．TV の Pr(> $|t|$) が 0.0034 なので，TV は SBP と関連していない．

ウ．TV の Pr(> $|t|$) が 0.0034 なので，TV により SBP の変動は十分説明される．

エ．Adjusted R-squared が 0.0007579 なので，TV により SBP の変動は十分説明されない．

オ．F-statistic に関する p 値が 0.003399 なので，TV により SBP の変動は十分説明される．

生物統計情報学 第2問 (選択問題)

既知のパラメータ $\lambda > 0$ と未知のパラメータ $\mu \geq 0$ によって定まる確率分布 $P_{\lambda,\mu}$ に従う確率変数 X を考える．確率分布 $P_{\lambda,\mu}$ の確率密度関数は次のように与えられている：

$$f(x) = \frac{\lambda}{2} \exp\left(-\lambda|x-\mu|\right) \qquad (x \in \mathbb{R})$$

このとき，以下の問いに答えよ．

(2-1) 確率変数 X の期待値と分散を求めよ．

(2-2) 確率分布 $P_{\lambda,\mu}$ の累積分布関数 $F(x)$ $(x \in \mathbb{R})$ を求めよ．

(2-3) 確率分布 $P_{\lambda,\mu}$ の上側 $100\alpha\%$ 点を求めよ．ただし，$0 < \alpha \leq 0.1$ とする．

(2-4) 未知パラメータ μ について，次の仮説検定を考える：

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu > 0$$

検定統計量を X，有意水準を α $(0 < \alpha \leq 0.1)$ として，第2種の過誤確率が最も小さくなるような棄却域を構成せよ．

(2-5) **(2-4)** の検定における検出力関数 $\beta_\lambda(\mu)$ $(\mu \geq 0)$ を求めよ．

生物統計情報学 第3問 (選択問題)

ある疾患を有する集団300名と当該疾患を有さない集団300名を対象とした臨床研究において，ある検査Aと検査Bの診断性能を調べた結果，次のような分割表データが得られた．

		疾患		
		あり ($D = 1$)	なし ($D = 0$)	合計
検査A	陽性 (+)	180	30	210
	陰性 (−)	120	270	390
	合計	300	300	600

		疾患		
		あり ($D = 1$)	なし ($D = 0$)	合計
検査B	陽性 (+)	270	120	390
	陰性 (−)	30	180	210
	合計	300	300	600

(3-1) 検査Aと検査Bの感度 $\Pr(+ \mid D = 1)$ と特異度 $\Pr(- \mid D = 0)$ の値を各々求めよ．

(3-2) **(3-1)** で求めた感度と特異度の値を踏まえ，検査Aと検査Bは各々，疾患のスクリーニングと確定診断のいずれの場面に有用と考えられるか．理由とともに説明せよ．

(3-3) 疾患の有病割合 $\Pr(D = 1)$ を p としたとき，検査Bの陽性適中度 $\Pr(D = 1 \mid +)$ および陰性適中度 $\Pr(D = 1 \mid -)$ を p の関数として表せ．

(3-4) 検査Bについて，陽性適中度と有病割合の関係および陰性適中度と有病割合 p の関係を表すグラフを，有病割合を横軸として各々描け．ただし，$p = 0.1, 0.4, 0.7$ の3点における各々の関数値をグラフ中に明示し，それらのグラフが単調増加関数か否かおよび上に凸な関数か否かについて，根拠を示すこと．

(3-5) この疾患の高リスク因子を有する集団に対して検査Bを適用することを考える．このとき，陽性適中度および陰性適中度の値を踏まえ，どのような問題が考えられるか述べよ．同様に，この疾患のリスク因子を有さない集団に対して検査Bを適用することを考える．このとき，陽性適中度および陰性適中度の値を踏まえ，どのような問題が考えられるか述べよ．