H28 海事統計学 期末試験（１／３）

以下、必要ならば $\sqrt{2} = 1.41$, $\sqrt{3} = 1.73$, $\sqrt{5} = 2.24$, $\sqrt{10} = 3.16$ として計算せよ。
答案には導出過程を記述すること。

問題１

以下のデータのメディアンとモードを求めよ。（5 点）

5	8	3	7	3	9	10	3	1	9

問題２

あるクラスの英語と数学の点数を解析したところ、数学の平均点が $\bar{x}$，数学の分散が s_x^2，英語の平均点が $\bar{y}$，英語の分散が s_y^2，英語と数学の間の共分散が s_{xy} であった。このとき数学と英語の点数データの相関係数を示せ。（5 点）

問題３

期待値 0，分散 1 の正規分布に従うサンプルを得るために、計算機上で使用可能な 0〜1 の一様分布に従う独立な複数のサンプルを利用することを考える。このとき以下の問に答えよ。

【問３–１】（5 点）
確率変数 x が $0 \leq x \leq 1$ の範囲において確率密度関数 $f(x) = 1$ で一様に分布するとき、この確率変数 x の期待値および分散を計算せよ。

【問３–２】（5 点）
期待値 μ，分散 σ^2 の任意の確率分布に従う独立な n 個のサンプル $x_1, x_2, \cdots x_n$ を足し合わせたときの分布（あるいはサンプルの平均値の分布）は、サンプル数 n が大きくなると正規分布に近づくことが知られている。この現象を示す定理は何と呼ばれるか答えよ。

【問３–３】（10 点）
$0 \leq x \leq 1$ の範囲において一様に分布する【問３–１】で用いた確率変数 x の独立な n 個のサンプル $x_1, x_2, \cdots x_n$ と、ある定数 A を用いて、新たな確率変数 $Z = (x_1 + x_2 + \cdots + x_n) - A = (\sum_{i=1}^{n} x_i) - A$ を <u>期待値 0，分散 1</u> の正規分布の近似として用いることを考える。そのためにはサンプル数 n および定数 A をどのように決めたら良いか答えよ。

【問３–４】（5 点）
期待値 3，分散 4 の正規分布に従う確率変数 x が $4 \leq x \leq 5$ の範囲の値となる確率を正規分布表を用いて求めよ。

問題４

あるメーカーで製造した１０個の電球の寿命を測定し、それぞれの時間を $x_1, x_2, \cdots, x_{10}$ と表して集計すると以下のようになった：

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^{10} x_i}{10} = 2593 \\ \sum_{i=1}^{10} (\bar{x} - x_i)^2 &= 57600\end{aligned}$$

これらのデータを用いて、このメーカーの電球の平均寿命を信頼係数 95 %で区間推定せよ。
ただし電球の寿命は正規分布に従うものと仮定する。（10 点）

問題５

ある造船所で同一種類の 27 個のブロック組立て作業に要した日数について、以下の表に示すデータが観測された。この表は、例えば「作業に要した日数＝１」において「観測度数＝７」となっているが、これは 27 個のブロックのうち 7 個のブロックが作業日数＝１だったことを表している。このデータについて以下の問いに答えよ。

作業に要した日数	1	2	3	4	5	6	7	8 日以上	計
観測度数	7	7	4	3	2	2	2	0	27 個

【問５–１】（5 点）
このブロック組立て作業に要する日数の平均値を求めよ。

【問５–２】（10 点）
このデータについて、作業日数を確率変数 x とした幾何分布であると仮定すると、表は 27 個分の標本 $x_1, x_2, \cdots, x_{27}$ について表している。このとき 1 日あたりに作業が完了する確率 p を表のデータから最尤推定により求めよ。

【問５–３】（5 点）
作業日数を確率変数 x とした幾何分布において 1 日あたりに作業が完了する確率が p のとき、作業日数 $x = 2$ となる確率を p を使った式で示せ。

【問５–４】（10 点）
表のデータが幾何分布に従っていると言えるかどうかについて、カイ 2 乗分布表を利用して有意水準 5 %で適合度検定せよ。

問題６

以下はある造船所で建造されたタンカーの主要目とその工事に伴う溶接長と塗料量の実績である。

船の番号	L	B	D	Cb	延べ溶接長	塗料合計
No.1	168	32.2	17.0	0.80	11.3	320
No.2	192	32.2	15.2	0.79	12.4	350
No.3	116	20.0	13.7	0.78	3.32	89.7
No.4	175	23.1	15.2	0.82	11.7	332
No.5	137	23.1	14.5	0.78	8.72	247
No.6	172	32.2	18.1	0.80	12.3	349
No.7	97	15.5	10.5	0.70	2.1	51.2
No.X	148	24.6	9.93	0.77	?	?

【問６－１】（15 点）
上記テーブルに示すようにデータが与えられたとき、工事予定の船（No.X）における延べ溶接長と塗料合計を多重回帰を用いて推定したい。
上記の表データの数字を用いて式を作り、計算手順を説明せよ。（数値計算はしなくて良い）

【問６－２】（10 点）
問１の推定をより精度良く行うため、以下のような項目 1,2,3 のような工夫を検討した。
このとき、明らかに推定計算上不都合が生じると考えられるものを全て挙げ、その理由を説明せよ。ただし単なる計算量の増加は、計算上の不都合とは考えないものとする。

1. 新しく番号 No.8 についてのデータをテーブルに追加し、問１–１と同様の推定方法で計算をやり直す。このとき、この No.8 は No.6 と同型船であるため、L, B, D, Cb の値は No.6 と完全に同一であるが、延べ溶接長と塗料合計は若干異なっている。
2. 各船の特徴量として L, B, D, Cb に加えて、船殻の体積を反映する新しい特徴量 $x_5 = L \cdot B \cdot D \cdot C_b$ を導入し、問６–１と同様の推定方法で計算をやり直す。
3. 各船の特徴量として L, B, D, Cb に加えて、船殻の長さを反映する新しい特徴量 $x_5 = L + B + D$ を導入し、問６–１と同様の推定方法で計算をやり直す。

海事統計学 平成28年度期末試験（2016年8月2日実施）解答

問題１

（5点）小さい順に並び替えると

1　3，3，3　5　7　8　9，9　10　　よって、メディアンは $(5+7)/2=6$　　モードは3

問題２（5点）

$$r=\frac{s_{xy}}{s_x s_y}$$

問題３

【問3–1】（5点）期待値を μ、分散を σ^2 とすると、

$$\mu=\int_0^1 x\,f(x)dx=\int_0^1 x dx=\left[\frac{x^2}{2}\right]_0^1=\frac{1}{2}$$

$$\sigma^2=\int_0^1(x-\mu)^2 f(x)dx=\int_0^1\left(x-\frac{1}{2}\right)^2dx=\left[\frac{1}{3}\left(x-\frac{1}{2}\right)^3\right]_0^1=\frac{1}{12}$$

【問3–2】（5点）中心値極限定理：母平均 μ，母分散 σ^2 の任意の母集団から独立に抽出された n 個の標本 $x_1,x_2,\ldots x_n$ の平均を $\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}$ とすると、標本数 n が大きくなると正規分布 $N(\mu,\frac{\sigma^2}{n})$ に近づく。
あるいは $\bar{x}$ を正規化して $z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ とすれば、確率変数 z は標本数 n が大きくなると $N(0,1)$ に近づく。

【問3–3】（10点）期待値 μ，分散 σ^2 の確率分布に従う確率変数 x の独立な n コのサンプルを用いた確率変数 $Y=x_1+x_2+\cdots+x_n=\sum_{i=1}^n x_i$ を定義する。このとき、独立な確率変数を足し合わせた値の期待値は、各確率変数の期待値の和に等しい。よって確率変数 Y の期待値は $n\mu$ となる。また、独立な確率変数を足し合わせた値の分散は、各確率変数の分散の和に等しい。よって確率変数 Y の分散は $n\sigma^2$ となる。ここで問題の確率変数 Z は確率変数 Y から定数 A を引いたものである。分散は定数 A の影響を受けないので、確率変数 Z の分散は確率変数 Y の分散に等しい。よって $n\sigma^2=1$
ここで一様分布の分散は【問3–1】より $\sigma^2=\frac{1}{12}$ である。よって $n=12$ となる。
また Z の期待値は Y の期待値から定数 A を差し引いた値に等しいので $n\mu-A=0$
一様分布の期待値は【問3–1】より $\mu=\frac{1}{2}$ である。よって $A=12\times\frac{1}{2}=6$
◎つまり 0〜1 の一様乱数を12個足して6を引くと平均値0分散1の正規分布の良好な近似サンプルが得られる。

【問3–4】（5点）期待値 μ，分散 σ^2 の正規分布に従う確率変数 x を標準化した確率変数

$$z=\frac{x-\mu}{\sigma}$$

は期待値0、分散1の標準正規分布 $N(0,1)$ に従う。よって $4\le x\le 5$ の範囲をとる確率は、$\frac{4-3}{2}\le z\le\frac{5-3}{2}$ の範囲をとる確率と等価である。標準正規分布表より、$0\le z\le 1$ の範囲の面積は 0.3413、$0\le z\le\frac{1}{2}$ の範囲の面積は 0.1915 なので、$0.3413-0.1915=0.1498$ が答えとなる。

問題４（10 点）

$\sum_{i=1}^{10}(\bar{x}-x_i)^2 = 57600 = 80\times 80\times 9$ より、標本不偏分散は 80
自由度９の t 分布表から $t_9(0.025) = 2.262$ であるから

$$-2.262 < \frac{2593-\mu}{80/\sqrt{10}} < 2.262$$

よって

$$2593 - 2.262\frac{80}{\sqrt{10}} < \mu < 2593 + 2.262\frac{80}{\sqrt{10}}$$
$$2535.7 < \mu < 2650.3$$

問題５

作業日数	1	2	3	4	5	6	7	8 日以上	計
観測度数 f	7	7	4	3	2	2	2	0	27
期待度数 f^*	9	6	4	$\frac{8}{3}$	$\frac{16}{9}$	$\frac{32}{27}$	$\frac{64}{81}$	$\frac{68}{81}$	27
分布確率	p	$p(1-p)$	$p(1-p)^2$	$p(1-p)^3$	$p(1-p)^4$	$p(1-p)^5$	$p(1-p)^6$	$1-p-\cdots-p(1-p)^6$	1

ただし $p = 1/3$

【問５–１】（5 点）作業日数の平均値：$(1\times 7+2\times 7+3\times 4+4\times 3+5\times 2+6\times 2+7\times 2)/27 \ = \ 3$

【問５–２】（10 点）1 日あたりに作業が完了する確率 p の最尤推定値は １／（作業日数の平均値）$= \frac{1}{3}$

【問５–３】（5 点）
作業日数 $x = 2$ となる確率＝ $p(1-p)$

【問５–４】（10 点）
帰無仮説：表のデータが幾何分布に従っている
対立仮説：幾何分布に従っていない
期待度数を計算すると、作業日数３以降のクラスにおいて期待度数が５を下回っているため、これらのクラスを統合すると、クラス数は３になる。検定を行う度数分布表は以下のとおり；

作業日数	1	2	3 日以上	計
観測度数 f	7	7	$4+3+2+2+2=13$	27
期待度数 f^*	9	6	12	27
分布確率	p	$p(1-p)$	$1-p-p(1-p)$	1

上記の度数分布表のあてはめの χ^2 統計量は以下のとおり：

$$\chi^2 = \frac{(7-9)^2}{9} + \frac{(7-6)^2}{6} + \frac{(13-12)^2}{12} = \frac{25}{36} = 0.694444\cdots$$

また幾何分布の未知母数 p をデータから最尤推定するため、自由度はクラス数３から推定した未知母数の個数１を差し引いて２となる。自由度２のカイ 2 乗分布の面積が $1-0.05=0.95$ になる領域は、カイ 2 乗分布表より 5.99 であるため、有意水準５％の棄却域は $5.99 < \chi^2$ の領域である。よって度数分布表のあてはめの χ^2 統計量は棄却域に入らないので、帰無仮説は棄却されない。よって表のデータは幾何分布に従っている（ことを否定できない）。

問題６

【問６－１】（15 点）特徴量 L, B, D, Cb を説明変数 x_1, x_2, x_3, x_4 で表し、延べ溶接長 y を以下の多重回帰で説明：

$$y \;=\; b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + e$$

ただし b_0, b_1, b_2, b_3, b_4 は回帰係数、e は確率変動（誤差）である。
ここで、表のデータを用いて、回帰係数 $b_0, b_1, \cdots, b_4$ を求める：

$$\boldsymbol{y} = \begin{bmatrix} 11.3 \\ 12.4 \\ 3.32 \\ 11.7 \\ 8.72 \\ 12.3 \\ 2.1 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 168 & 32.2 & 17.0 & 0.80 \\ 1 & 192 & 32.2 & 15.2 & 0.79 \\ 1 & 116 & 20.0 & 13.7 & 0.78 \\ 1 & 175 & 23.1 & 15.2 & 0.82 \\ 1 & 137 & 23.1 & 14.5 & 0.78 \\ 1 & 172 & 32.2 & 18.1 & 0.80 \\ 1 & 97 & 15.5 & 10.5 & 0.70 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix}$$

【注意１】行列 X には、推定の対象である「工事予定の船 No.X」の要目データを含めない

【注意２】行列 X の１列目の要素はすべて 1
と表すと、$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$ とした場合の誤差ベクトル $\boldsymbol{e}$ の平方和を最小にする回帰係数 $\hat{\boldsymbol{b}}$ は以下の式で与えられる：

$$\hat{\boldsymbol{b}} = \left(\boldsymbol{X}^{Trans}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{Trans}\boldsymbol{y}$$

工事予定の船 No.X の要目 L, B, D, Cb を $x_1 = 148, x_2 = 24.6, x_3 = 9.93, x_4 = 0.77$ へ代入し、上で求めた回帰係数 $\hat{\boldsymbol{b}}$ を使って $y \;=\; \hat{b_0} + \hat{b_1} x_1 + \hat{b_2} x_2 + \hat{b_3} x_3 + \hat{b_4} x_4$ より工事予定の船 No.X の延べ溶接長の推定値 y を得る。

【注意３】推定時には誤差 e の項をゼロで計算する。

また、塗料合計については上記の行列のうち $\boldsymbol{y}$ だけを以下のように設定して同様の計算にて推定する：

$$\boldsymbol{y} = \begin{bmatrix} 320 \\ 350 \\ 89.7 \\ 332 \\ 247 \\ 349 \\ 51.2 \end{bmatrix}$$

【問６－２】（10 点）説明変数 $x_1, x_2, \cdots x_k$ について、任意の定数 $a_0, a_1, \cdots a_k$ を用いて関係式 $a_0 + a_1 x_1 + a_2 x_2 + \cdots a_k x_k = 0$ に近い関係が成り立つとき、データは強い多重共線関係（多重共線性）があるという。このとき回帰係数ベクトルを求める逆行列計算が不安定になり、無意味な解が出やすい。

ここで、問題文のうち、明らかに 3 番目の方法は上記の関係式を生み出してしまい、回帰係数ベクトルを求める逆行列計算ができないという不都合が生じる。

多重回帰は最小２乗法であり、データの重複は問題ないので、1 番目の方法は問題ない。
また２番目の方法は、新しい説明変数を既存の説明変数から生成しているが、非線形関数になっているので多重共線性は生じず、問題は無い。

本試験問題および解答に疑問がある場合は，8 月 9 日（火）午後４時までに，
W２号館６３４号室の木村まで申し出ること。

平成 28 年度 海事統計学 解答用紙

学籍番号＿＿＿＿＿＿＿＿氏名＿＿＿＿＿＿＿＿＿＿

紙面が足りない場合は裏面も使用すること。