海事統計学 定期試験問題（１／３）

答案には導出過程を記述すること。

問題１

３枚のコインを同時に投げ、表になるコインの枚数を確率変数 x とする。このとき、この確率変数 x の期待値と分散を求めよ。(10 点）

問題２

期待値 4, 分散 16 の正規分布に従う確率変数 x が $5 \leq x \leq 6$ の範囲の値となる確率を求めよ。(10 点）

問題３

互いに独立な確率変数 A, B, C が期待値 μ および分散 σ^2 に従うとき、これらの確率変数の和で作られた新しい確率変数 $Z = A + B + 2C$ の期待値および分散の値を μ および σ を用いて表せ。（10 点）

問題４

母比率 p を n 個のデータから得られた標本比率を用いて信頼係数 95 %で区間推定することを考える。信頼区間の幅を 0.1 以下とするには、標本の大きさ n をいくら以上にしたら良いか答えよ。（15 点）

海事統計学 定期試験問題（２／３）

問題５

以下の表は、ある類似船型の貨物船 12 隻において 1 年間にバルブの故障が発生した件数を表している。
このとき以下の問いに答えよ。

船の番号	1	2	3	4	5	6	7	8	9	10	11	12
年間故障発生件数	0	1	0	0	4	1	0	2	2	0	1	1

【問５–１】（5 点）
１隻あたり１年間で発生するバルブ故障の平均件数を計算せよ。

【問５–２】（5 点）
年間に発生する故障の発生件数 x の確率 $P(x)$ をポアソン分布 $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ のモデルで説明したい。
船の番号 i における年間故障発生件数を x_i と表す。このときポアソン分布によるモデルの対数尤度を x_i を用いて式で示せ。

【問５–３】（5 点）
問４–２の対数尤度を最大化する λ はどのような値か？　x_i を用いて式で示せ。

【問５–４】（5 点）
表に示されたデータを最も尤もらしく説明するポアソン分布モデルの λ はどのような値か？
表のデータから計算して示せ。

【問５–５】（10 点）
表に示されたデータがポアソン分布に従っているといえるかどうか検定したい。どのような検定方法が考えられるか検討せよ。また検定方法が使えないなら、その理由を述べよ。

【問５–６】（5 点）
ポアソン分布と指数分布の関係について説明せよ。

海事統計学 定期試験問題（３／３）

問題６

以下は 1980 年代の自動車の要目と燃費についてのデータである。

車種＼要目	x_1	x_2	x_3	x_4	x_5	x_6	10 モード 走行 (km/L)
クラウン	1.360	4.778	2.4251	125	17.5	8.8	8.7
マーク II	1.245	4.100	2.4082	125	17.5	8.8	9.5
カムリ	1.070	3.214	2.3575	120	17.6	8.7	10.6
ソアラ	1.235	4.100	2.3052	125	17.5	8.8	9.2
セド リック	1.420	4.625	2.4251	130	17.5	9.5	8.9
ローレル	1.175	3.889	2.3660	125	17.0	9.1	9.2
スカイライン	1.175	4.111	2.3198	125	17.0	9.1	9.2
レパード	1.220	3.900	2.2899	125	17.0	9.1	9.4
カペラ	1.030	3.450	2.3829	120	17.0	8.6	10.2
ギャラン Σ	1.180	3.665	2.3645	110	16.7	8.5	10.6
ルーチェ	1.150	3.909	2.3829	120	17.0	8.6	?

ただし要目 x_i については、x_1:車体重量 (トン)，x_2：減速比，x_3：幅×高さ (m^2)，x_4：最大出力 (ps)，x_5：最大トルク (kgm)，x_6：圧縮比

【問６－１】(15 点）上記のデータのうち、ルーチェの 10 モード 走行の値が不明であるので、他の車のデータを利用して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、ルーチェの燃費推定の計算手順を説明せよ。

【問６－２】(5 点）多重回帰による推定において問題となる多重共線性とは何か、どのようなデータを扱うと生じるかについて説明せよ。

海事統計学 平成26年度定期試験問題（2014年7月28日実施）解答

問題１解答

二項分布の期待値と分散より $E\{x\} = np = 3 \times 0.5 = \frac{3}{2}$,　　$\mathrm{Var}\{x\} = np(1-p) = 3 \times 0.5 \times (1-0.5) = \frac{3}{4}$（10点）

問題２ 解答

【正規分布の標準化と正規分布表の読み取り】（10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。よって期待値4, 分散16の正規分布に従う確率変数 x が $5 \leq x \leq 6$ の範囲をとる確率は、標準正規分布 $N(0,1)$ が $\frac{5-4}{4} \leq z \leq \frac{6-4}{4}$ の範囲をとる確率と等価である。標準正規分布表より、$0 \leq z \leq 1/4$ の範囲の面積は 0.0987、$0 \leq z \leq \frac{1}{2}$ の範囲の面積は 0.1915 なので、$0.1915 - 0.0987 = 0.0928$ が答えとなる。

問題３ 解答

期待値 4μ,　　分散 $6\sigma^2$　　（10点）

問題４解答

【区間推定】（15点）
標本 n が大標本と仮定すると、標本比率は正規分布で近似できる。標準正規分布における 95 %の信頼区間は ± 1.96 であるので、

$$2 \times 1.96 \sqrt{\frac{p(1-p)}{n}} \leq 0.1$$

ここで $0 \leq p \leq 1$ なので $p(1-p)$ の最大値は 0.25．よって

$$2 \times 1.96 \sqrt{\frac{0.25}{n}} \leq 0.1$$

式変形すると

$$n \geq 384.16$$

よって標本の大きさ $n \geq 385$　（15点）

問題５解答

【問５–１】（５点）１件／年

【問５–２】（５点）

$$\sum_{i=1}^{12} \ln P(x_i) = \sum_{i=1}^{12} \ln \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

【問５–３】（５点）

$$\frac{\partial}{\partial \lambda} \sum_{i=1}^{12} \ln \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = 0$$

この方程式を λ について解くと

$$\lambda = \frac{\sum_{i=1}^{12} x_i}{12}$$

【問５–４】（５点）
問５－３より、標本平均に等しいので $\lambda = 1$

【問５–５】（10 点）
正規分布でも比率の分布でもないので、それらの平均値や分散を検定する方法は使えない。
カイ２乗分布を使ったあてはまりの検定は、期待度数を５以上にしなければならないが、それができないので使えない。
また K-S 検定はデータ数が 20 より大きくないと使えない。よって講義で扱った検定方法はどれも適用できない。

【問５–６】（５点）
ポアソン分布における発生時間間隔は、指数分布に従う。事象の発生時間間隔を x としたとき、指数分布の確率密度関数は $f(x) = \lambda e^{-\lambda x}$ で表される。このときの指数分布のパラメータ λ の値はポアソン分布のパラメータ λ と同一になる。

問題６解答

【問６−１】(15 点)
10 モード走行 (km/L) を y とし、各要目の値 $x_1, x_2, x_3, x_4, x_5, x_6$ を変数（説明変数）として以下の多重回帰によって説明する：

$$y \;=\; b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + e$$

ただし $b_0, b_1, b_2, b_3, b_4, b_5, b_6$ 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 $b_0, b_1, \cdots, b_6$ を求める。行と列を間違えないよう注意しながら表のデータを行列に変換すると、以下のようになる：

$$\boldsymbol{y} = \begin{bmatrix} 8.7 \\ 9.5 \\ 10.6 \\ 9.2 \\ 8.9 \\ 9.2 \\ 9.2 \\ 9.4 \\ 10.2 \\ 10.6 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1.360 & 4.778 & 2.4251 & 125 & 17.5 & 8.8 \\ 1 & 1.245 & 4.100 & 2.4082 & 125 & 17.5 & 8.8 \\ 1 & 1.070 & 3.214 & 2.3575 & 120 & 17.6 & 8.7 \\ 1 & 1.235 & 4.100 & 2.3052 & 125 & 17.5 & 8.8 \\ 1 & 1.420 & 4.625 & 2.4251 & 130 & 17.5 & 9.5 \\ 1 & 1.175 & 3.889 & 2.3660 & 125 & 17.0 & 9.1 \\ 1 & 1.175 & 4.111 & 2.3198 & 125 & 17.0 & 9.1 \\ 1 & 1.220 & 3.900 & 2.2899 & 125 & 17.0 & 9.1 \\ 1 & 1.030 & 3.450 & 2.3829 & 120 & 17.0 & 8.6 \\ 1 & 1.180 & 3.665 & 2.3645 & 110 & 16.7 & 8.5 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$

【注意１】行列にルーチェのデータを含めないこと　【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：

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

ルーチェの要目を $x_1 = 1.150, x_2 = 3.909, x_3 = 2.3829, x_4 = 120, x_5 = 17.0, x_6 = 8.6$ へ代入し、上で求めた回帰係数 $\hat{\boldsymbol{b}}$ を使って $y \;=\; \hat{b_0} + \hat{b_1}x_1 + \hat{b_2}x_2 + \hat{b_3}x_3 + \hat{b_4}x_4 + \hat{b_5}x_5 + \hat{b_6}x_6$ よりルーチェの 10 モード走行 (km/L) の推定値 y を得る。
【注意３】この推定時の式には誤差 e は入れないこと

【問６−２】(5 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 $a_0 + a_1x_1 + a_2x_2 + \cdots a_kx_k = 0$ 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。

本試験問題および解答に疑問がある場合は，８月２５日（月）午後５時までに，
Ｗ２号館６３４号室の木村まで申し出ること。

2014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。