確率統計の演習問題

目次

確率統計の演習問題です。原本は「統計学演習」村上正康、安田正實；培風館（ISBN 978-4-563-00870-3） のテキストです。この中から選んでいますし、変更や追加もしています。解答は原本を参照のこと。

1 データの整理

問 1.1. ある日、図書館で本を借りた人の年齢を調べたところ、

年齢	*8-12*	*13-20*	*21-60*	*61-64*
人数	9	22	35	14

(1) このデータをヒストグラムで表せ。 (2) この日の集計結果における年齢の平均と標準偏差をもとめよ。

問 1.2. つぎの *8* 個のデータ*:* $6, 8, 8, 10, 21, x, y, z$、ただし *(*x, y, z は未知*)* は、平均が 7 で、モード（最頻値）が 6 であるという。このときメディアン（中央値、中位数）はいくつか？

問 1.3. あるデータを集計したところ、度数分布表はつぎの結果 (i) となった。

(1) この (i) の表から、その標本平均と標本分散、標準偏差を求めよ。
(2) また (ii) の表における場合の計算を (i) の結果から計算せよ。

(i)

値	*110*	*120*	*130*	*140*	*150*	計
度数	2	6	3	7	2	20

(ii)

値	*10*	*20*	*30*	*40*	*50*	計
度数	2	6	3	7	2	20

問 1.4. 2 つのグループ *A,B* のうち、A は大きさ 4、平均 5*,* 分散 2 であり、B は大きさ 6*,* 平均 7*,* 分散 3 であった。この *2* つを合体させたとき、平均と分散をもとめよ。

2　確率と確率分布

問 2.1. 2 つの事象 A, B の確率が $P(A \cup B) = \dfrac{7}{8}$, $P(A \cap B) = \dfrac{1}{4}$, $P(\overline{A}) = \dfrac{5}{8}$, と与えられている。つぎの確率をもとめよ。　(i)　$P(A)$　(ii)　$P(\overline{B})$　(iii)　$P(A \cap B)$　(iv)　$P(A \cup \overline{B})$

問 2.2. 事象 A , B は互いに排反 *(*同時に起ることはない*)* 事象で $P(A) = 0.2$, $P(B) = 0.8$ で、さらにある事象 C の条件つき確率が $P(C \,|\, A) = 0.4$ と $P(C \,|\, B) = 0.5$ で与えられている。$P(A \,|\, C)$ と $P(B \,|\, C)$ をもとめよ。答えは分数のままでもよい。

問 2.3. 全事象 $S = \{1, 2, 3, 4\}$ とし、それぞれ 4 個の点に対して、同じ確率 $\frac{1}{4}$ ずつをもつとする。3 つの事象 $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$ において、(i) $P(AB) = P(A)P(B)$, $P(AC) = P(A)P(C)$, $P(BC) = P(B)P(C)$ を示せ。　(ii) しかし、3 つが同時に起こる積事象の確率 $A \cap B \cap C = ABC$ はそれぞれの確率の積にはならない。すなわち

$$P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

問 2.4. 確率変数 X の平均（期待値） が $E(X) = m$, 分散が $V(X) = s^2$ であるという。つぎの確率変数 (a) $Y = 3X + 2$, (b) $Z = X - m$, (c) $W = \dfrac{X + b}{a}$ （ここで a, b は定数*)* について、それぞれの平均と分散をもとめよ。

問 2.5. 離散型確率変数 (X, Y) の結合分布 *:* $f(x, y) = P(X = x, Y = y)$ がつぎで与えられている。

$$f(1,1) = 1/8,\ f(1,2) = 1/4,\ f(2,1) = 1/8, f(2,3) = 3/8,\ f(2,4) = 1/8,$$

その他の値では 0。このとき、(1) 和 $X + Y$ の平均をもとめよ。(2) 積 XY の平均をもとめよ。(3) この 2 つの確率変数独立となるかどうかを調べよ。

問 2.6. 箱の中に 4 個の赤球 と 3 個 の青球がある。 この箱から, 　「非復元抽出」で青球が出るまで球のとり出しを続ける。X を青球が出るまでの球のとり出し回数とする。
(1)　X の確率分布を求めよ。　(2)　X の平均 $E(X)$ と 分散 $V(X)$ を求めよ。

3　2 項分布と正規分布

正規分布の確率を計算するには、原本テキストの末尾部分にある「統計分布表」か、「表計算ソフトの命令」をつかえば簡単に求められる。

問 3.1. *2* 項分布 $X \sim Binom(n,p)$ において、 $p = \dfrac{4}{5}$, $P(X=4) = 5\,P(X=5)$ のとき、平均 $E(X)$ をもとめよ。

問 3.2. ある集団には左利きの人が 10% いるという。この中から n を選んだとき、そのうちに少なくとも一人以上左利きの人が含まれる確率を 0.95 以上にするには、n の値はいくつにしなければならないか。

問 3.3. *2* 項分布 $X \sim Binom(8, \frac{1}{2})$ において、 つぎをもとめよ。

(1) 平均 $m = E(X)$ および分散 $s^2 = V(X) = E(X-m)^2$ (2) 確率 $P(|X-m| \geq 2s)$ の値

問 3.4. $Z \sim N(0,1)$ のとき、正規分布表（*WEB* 配布の資料あるいは演習書の *p.203, 204*）または表計算ソフト (=norm.s.dist(引数), =norm.s.inv(引数) など) から、つぎの確率を計算せよ。

(1) $P(\,Z < 1\,)$ (2) $P(\,Z < 1.24\,)$ (3) $P(\,-1 \leq Z < 1.24\,)$

(4) $P(\,2Z+3 < 4\,)$ (5) $P(\,|\,Z-1\,| < 0.24\,)$

問 3.5. コインを 400 回投げるとき、表の出た回数が 180 回から 210 回の間となる確率を正規近似をもちいて計算せよ。

4 標本分布

問 4.1. 2 つの独立な正規分布の一次結合のつくる分布を調べる。X, Y はそれぞれ標準正規分布 $N(0,1)$ にしたがうとき、

(1) $2X+Y$ の平均 (2) $2X+Y$ の分散 (3) 確率 $P(2X+Y>3)$ (4) 確率 $P(1<2X+Y<3)$

問 4.2. *5* 個の正規分布にしたがう $X_i \sim N(2,1)$, $i = 1, 2, \cdot, 5$ に対して、平均 $Y = \dfrac{1}{5}(X_1 + \cdots + X_5)$ をつくる。

(1) Y の平均と分散をもとめよ。 (2) $P(2<Y<4)$ をもとめよ。

問 4.3. 2 つの正規分布にしたがう $X \sim N(5,9)$, $Y \sim N(7,16)$ は独立であるとき、

(1) $X-Y$ の分布は何か、 (2) $3X+Y$ の分布は何か (3) $\overline{X}_{25} - \overline{Y}_{16}$ の分布は何か

ここで $\overline{X}_{25}, \overline{Y}_{16}$ はそれぞれの分布からの大きさ 25, 16 の標本平均とする。

問 4.4. 平均 μ, 分散 σ^2 の正規母集団から大きさ $n = 4$ の標本平均を $\overline{X}$ とする。

(1) 確率 $P(|\overline{X} - \mu| < \sigma)$ の値をもとめよ。

(2) 確率 $P(|\overline{X} - \mu| > 2\sigma)$ の値をもとめよ。

5 信頼区間

問 5.1. 正規母集団 $N(\mu, \sigma^2)$ からの n 個の標本を $X_1, X_2, \cdots, X_n$ とおく。標本平均 $\overline{X}_n$ から母平均 μ の推定をおこなう。

(1) 分散が既知のばあいに、信頼係数 $100(1-\alpha)\%$ の信頼区間の記述せよ。

(2) 分散が未知の場合で、大標本では分散の部分にはどう修正したらよいか？

問 5.2. ある学校の生徒の集団から、40 人を無作為に選んでテレビの視聴時間を調べた。そのデータの平均は 18.2 時間、標準偏差は 5.4 時間であった。このときこの学校の生徒集団に対してテレビ視聴時間の分布を正規分布として、その平均の μ を推定する。

(1) 95% 信頼区間 (2) 98% 信頼区間

をもとめよ。

問 5.3. 大標本のばあいの平均の差に対する信頼区間を調べる。

A 社の電球 80 個の寿命時間を調べたところ、平均 1070 時間、分散 472 で、一方 B 社の電球 60 個の寿命時間を調べたところ、平均 1042 時間、分散 366 であった。２つの集団分布の平均差 $\mu_A - \mu_B$ を推定する。２つの集団の平均差 $\mu_A - \mu_B$ の公式から

(1) 信頼係数 95% のとき、 (2) 信頼係数 90% のとき、

をもとめよ。

6 検定

問 6.1. ある箱のなかにある球の個数は、仮説 A : 赤球３個と黒球７個、または 仮説 B : 赤球６個と黒球４個のいづれかである。このとき、$\left\{ \begin{array}{l} \text{帰無仮説 } H_0 : \text{仮説 } A \text{ が正しい} \\ \text{対立仮説 } H_1 : \text{仮説 } B \text{ が正しい} \end{array} \right\}$ としたとき、標本を２個選んだ結果から判断する。非復元抽出をおこない、その結果が少なくとも *1* 個が黒であれば H_0 を採択し、それ以外は H_0 を棄却する。このような検定であれば、第１種の過誤、第２種の過誤の確率はいくつかになるか？

問 6.2. 正規母集団 $N(\mu, \sigma^2)$ の平均値 μ に対する検定をつぎの (1) ∼ (4) の場合に決定せよ。ただし n, n_1, n_2 は各集団における標本サイズ、$\overline{X}, \overline{X_1}, \overline{X_2}$ は各変数に対する標本平均、σ^2 は母分散、s, s_1, s_2 は標本標準偏差、α は有意水準をあらわす。

(1) $n = 100,\ \overline{X} = 38,\ \sigma = 25,\ \alpha = 0.05$ のとき、帰無仮説 $H_0 : \mu = 40$, 対立仮説 $H_1 : \mu \neq 40$
(2) $n = 16,\ \overline{X} = 240,\ \sigma = 20,\ \alpha = 0.05$ のとき、帰無仮説 $H_0 : \mu = 250$, 対立仮説 $H_1 : \mu < 250$
(3) $n = 14,\ \overline{X} = 51.52,\ s = 2.13,\ \alpha = 0.01$ のとき、帰無仮説 $H_0 : \mu = 50$, 対立仮説 $H_1 : \mu > 50$
(4) $n_1 = 50,\ n_2 = 40,\ \overline{X_1} = 22.31, \overline{X_2} = 21.54,\ s_1 = 3.8, s_2 = 3.2,\ \alpha = 0.05$ のとき、帰無仮説 $H_0 : \mu_1 = \mu_2$, 対立仮説 $H_1 : \mu_1 \neq \mu_2$

問 6.3. つぎの母集団比率の検定においては大標本と考え、正規近似を当てはめて検定せよ。

(1) $n = 120,\ \overline{X} = 90,\ \alpha = 0.05$ のとき、帰無仮説 $H_0 : p = 0.7$, 対立仮説 $H_1 : p \neq 0.7$
(2) 2集団の比較 : $n_1 = 420,\ \overline{X_1} = 118,\ n_2 = 530,\ \overline{X_2} = 121$ について、 $\alpha = 0.05$ のとき、帰無仮説 $H_0 : p_1 = p_2$, 対立仮説 $H_1 : p_1 > p_2$

問 6.4. あるクラス別の科目の成績評価 $\{A,B,C,D\}$ が つぎの結果であった。クラス間に成績の優劣が認められるかどうか、有意水準 *1%* で検定せよ。

	クラス 1	クラス 2	クラス 3
評価 A	8	12	6
評価 B	25	32	10
評価 C	10	9	14
評価 D	7	5	12
計	50	58	42

7　相関と回帰

問 7.1. つぎの場合について、2つの変量間の相関係数を求めよ。また散布図を描き、位置関係を確認せよ。

(a) 5個のデータ (X, Y):

X :	10	20	30	40	50
Y :	2	4	6	8	10

(b) 8個のデータ (Z, W):

Z :	1	1	1	0	0	−1	−1	−1
W :	1	0	−1	1	−1	1	0	−1

(c) 7個のデータ (U, V):

U :	−3	−2	−1	0	1	2	3
V :	8	3	0	−1	0	3	8

問 7.2. (i) X, Y の共分散を $s_{x,y}$ とするとき、この変量を $Z = aX + b,\ W = cY + d$ と変換したならば、Z, W の共分散 $s_{z,w}$ ともとの共分散 $s_{x,y}$ のあいだに成り立つ関係式を示せ。(ii) 2変量 X, Y を一次変換して変量 $Z = aX + b,\ W = cY + d$ と定めた相関係数でも同じ値となることを示せ。

問 7.3. 10 人の生徒について, 身長 (X) とひじの長さ (Y) を測定したら、次のデータを得た。単位はｃｍ）

X :	152	145	164	153	132	148	149	138	142	150
Y :	22	21	25	23	20	21	22	21	22	24

このとき、(i) *2* 変量の間の相関係数を求めよ。(ii)　変量 Y の変量 X への回帰直線をもとめよ。(iii) この回帰直線から、身長が 150 ｃｍ の生徒に対するひじの長さはどのくらいか？

問 7.4. n 個の標本が *2* つ の変量 x, y に ついてそれぞれ順位がつけられているとき, これら *2* 組の順位 $(x_1, x_2, \cdots, x_n)$ と $(y_1, y_2, \cdots, y_n)$ の間の相関係数（スピアマンの順位相関係数） r_s は

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)}$$

となることを示せ。