

2009 年春学期 統計の手法レポート No.1

解説と解答例

湊野 昌 (Sakaé Fuchino)

fuchino@isc.chubu.ac.jp

(2009 年 06 月 04 日)

以下は、2009 年春学期に中部大学で開講している「統計の手法」(木曜 5~6 時限)の第 1 回目のレポートの問題とその解説 / 解答例です。このテキストは

<http://pauli.isc.chubu.ac.jp/~fuchino/chubu/statistics-ss09-report01.pdf>

としてダウンロードできます。

解答例は十分に注意をはらって作成しているつもりですが、もし何かの誤りや問題点などを発見したときにはお知らせください。

1. 次の表は、あるクラスで行なったテスト（100点満点）の得点のデータです：

57	71	62	65	62	88	95	66	100	71	73	79	45	32	78	75	86	67	50	56	100	73	
79	35	75	50	56	88	75	90	63	82	63	82	63	72	87	49	48	69	45	90	73	82	78
72	83	55	80	54																		

- (a) このデータのメディアン，平均値，分散，標準偏差を求めてください．
- (b) このデータの平均値を中心としてプラスマイナス $2 \times$ 標準偏差の範囲の点を得た学生の全体に対する割合を計算してください．この値がチェビシェフの定理（教科書の p.30）と矛盾しないことを確認してください．
- (c) 10点台，20点台，30点台，…，90点台（ただし100点は90点台に含めることにする）を階級とするこの成績の度数分布表とヒストグラムを作成してください．
- (d) この試験を実施した先生は，それぞれの素点 x を $40 + \frac{6}{10}x$ で計算される点数に変換して成績簿に記入することにしました．成績簿上のデータのメディアン，平均値，分散，標準偏差は何になるでしょうか？ ヒント：教科書 p.26 の「ポイント」を応用できます（ただしメディアンについては，ここに出ていないので，このポイントの証明と類似の考察を自分でしてみる必要があります）．

メディアンを求めるためにデータを小さい順にならべなおすと：

3	32	35	45	45	48	49	50	50	54	55	56	56	57	62	62	63	63	63	65	66	67	69	71	71	72	72	73	73	75	75	75	78	78	79	79	80	82	82	82	83	86	87	88	88	90	90	95	100	100
---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

となる．データのサイズは50だから，このうちの25番目と26番目を見ると，72と72である．したがって，このデータのメディアンは72（点）であることがわかる．

平均値（平均点）は，

$$\frac{1}{50} (3 + 32 + 35 + \dots + 100) = 68.38 \approx 68 \text{ (点)}$$

分散は，

$$\frac{1}{50} ((3 - 68.38)^2 + (32 - 68.38)^2 + (35 - 68.38)^2 + \dots + (100 - 68.38)^2) = 337.6356$$

標準偏差は，

$$\sqrt{337.6356} = 18.374863264797373 \dots \approx 18 \text{ (点)}$$

(b): (a) から 平均点 $- 2 \times$ 標準偏差 $= 32$, 平均点 $+ 2 \times$ 標準偏差 $= 104$ となるから，この範囲の点を得た学生の数は，49人となる．チェビシェフの定理は，この範囲にいる学生の数は $(1 - \frac{1}{2^2}) \cdot 50 = 37.5$ 人以上であることを主張しているのだから，この数字はチェビシェフの定理と矛盾しない．

(c): 素点 x 点を $y = 30 + \frac{7}{10}x$ 点に変換するとき，教科書 p.26 のポイントにより，上のデータから変換によって得られたデータの平均値，分散，標準偏差はそれぞれ， $30 + \frac{7}{10} \times 68 = 77.6 \approx 78$ (点)， $(\frac{7}{10})^2 \times 337.6356 = 165.441444 \dots$ ， $\sqrt{165.441444 \dots} = 12.862404285358162 \dots \approx 13$ (点) となる．また，メディアンは，上の小さい順にならべかえた表で，値を $x \mapsto y = 30 + \frac{7}{10}x$ と変換したときの25番目と26番目の値の平均だから，その値は $30 + \frac{7}{10} \times 72 = 80.4$ 点となることがわかる．

2. 次は，ある変数 x と y の測定データです．

x	6	7	4	5	8
y	5	6	4	4	6

- (a) x と y の相関図を描いてください .
 (b) x と y の相関係数を求めてください .
 (c) このデータの回帰直線をあらわす式を求め , (a) で作成した相関図にこの回帰直線を記入してください .

注意: この問題はあくまで演習用です . 一般には , サイズが 5 くらいの小さな多変量データの相関からはあまり意味のある結論は引き出せません .

(a): 略 .

(b): x と y の平均はそれぞれ $\bar{x} = \frac{6+7+4+5+8}{5} = 6$, $\bar{y} = \frac{5+6+4+4+6}{5} = 5$ だから , x と y の標準偏差は , それぞれ $\sigma(x) = \sqrt{\frac{(6-6)^2+(7-6)^2+(4-6)^2+(5-6)^2+(8-6)^2}{5}} = \sqrt{2}$, $\sigma(y) = \sqrt{\frac{(5-5)^2+(6-5)^2+(4-5)^2+(4-5)^2+(6-5)^2}{5}} = \sqrt{0.8}$ となる . 教科書 37 ページの相関係数の式に , これら値 , および $N = 5$ と データの各値を代入すると ,

$$r(x, y) = \frac{1}{5} \frac{1}{\sqrt{2}\sqrt{0.8}} ((6-6)(5-5) + (7-6)(6-5) + (4-6)(4-5) + \dots) \approx 0.95$$

(c):

$$C(x, y) = \frac{1}{5} ((6-6)(5-5) + (7-6)(6-5) + (4-6)(4-5) + \dots) = \frac{6}{5}$$

だから , 教科書 41 ページによりこのデータでの y の x への回帰直線は

$$y = \frac{6}{5} \cdot \frac{1}{2}(x-6) + 5$$

つまり $y = \frac{3}{5}x + \frac{7}{5}$ で表される .

3. さいころを 2 ところがしたとき , 出た目の積を返す確率変数を X とします .

- (a) $P(X = 1)$, $P(X = 2)$, ... , は何になるか考えてください .
 (b) $P(3 \leq X \leq 8)$ は何になるでしょうか?
 (c) $E(X)$, $V(X)$, $\sigma(X)$ を求めてください .
 (d) $V(3X + 2)$ と $\sigma(3X + 2)$ を求めてください .
 (e) $E(2X^2 - X + 3)$ を求めてください .

(a): 教科書 48 ページの下の表と同様な表をここでの X に対して作ると ,

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

となるから , たとえば , X の値が 1 となるのは $6 \times 6 = 36$ 個の組合せのうち 1×1 の 1 つなので , $P(X = 1) = \frac{1}{36}$ となる . また , X の値が 2 となるのは 1×2 と 2×1 の 2 つの組合せとなるので , $P(X = 2) = \frac{2}{36} = \frac{1}{18}$ となる . 同様にして ,

$$P(X = 3) = \frac{2}{36} = \frac{1}{18}, P(X = 4) = \frac{3}{36} = \frac{1}{12}, P(X = 5) = \frac{4}{36} = \frac{1}{9},$$

$$P(X = 6) = \frac{4}{36} = \frac{1}{9}, P(X = 7) = 0, P(X = 8) = \frac{2}{36} = \frac{1}{18}, P(X = 9) = \frac{1}{36}, P(X = 10) = \frac{2}{36} = \frac{1}{18},$$

$$P(X = 11) = 0, P(X = 12) = \frac{4}{36} = \frac{1}{9}, P(X = 13) = 0, P(X = 14) = 0, P(X = 15) = \frac{2}{36} = \frac{1}{18},$$

$P(X = 16) = \frac{1}{36}, P(X = 17) = 0, P(X = 18) = \frac{2}{36} = \frac{1}{18}, P(X = 19) = 0, P(X = 20) = \frac{2}{36} = \frac{1}{18},$
 $P(X = 21) = P(X = 22) = P(X = 23) = 0, P(X = 24) = \frac{2}{36} = \frac{1}{18}, P(X = 25) = \frac{1}{36}, P(X = 26) =$
 $P(X = 27) = P(X = 28) = P(X = 29) = 0, P(X = 30) = \frac{2}{36} = \frac{1}{18}, P(X = 31) = \dots = P(X = 35) =$
 $0, P(X = 36) = \frac{1}{36}.$

また $k > 36$ に対しては $P(X = k) = 0$ である .

(b): $P(2 \leq X \leq 8) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$
 だから , (a) から , $P(2 \leq X \leq 8) = \frac{2+2+3+2+2+4+0+2}{36} = \frac{17}{36}$ である .

(c):

$$E(X) = 1 \cdot P(X = 1) + \dots + 36 \cdot P(X = 36) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{2}{36} + \dots + 36 \cdot \frac{1}{36} = 49/4 = 12.25$$

$$V(X) = (1 - E(X))^2 \cdot P(X = 1) + \dots + (36 - E(X))^2 \cdot P(X = 36) = 11515/144 \approx 79.97$$

$$\sigma(X) = \sqrt{V(X)} \approx 8.9$$

(d): 教科書 55 ページの囲みの中の (3) と上の (c) から

$$V(3X + 2) = 3^2 V(X) \approx 719.69$$

$$\sigma(3X + 2) = \sqrt{V(3X + 2)} \approx 26.8$$

(e): 教科書 55 ページの囲みの中の (1) と教科書 65 ページの囲みの中の (1) から

$$E(2X^2 - X + 3) = 2E(X^2) - E(X) + 3$$

したがって , 教科書 55 ページの囲みの中の (2) から ,

$$E(2X^2 - X + 3) = 2(V(X) + (E(X))^2) - E(X) + 3$$

である . この式に上での計算結果を代入して , $E(2X^2 - X + 3) \approx 888.98$