

統計の手法レポート No.2

解説と解答例

湊野 昌 (Sakaé Fuchino)

fuchino@isc.chubu.ac.jp

(2007年07月01日)

以下は、2007年春学期に中部大学で開講している「統計の手法」(火曜7~8時限のクラスと水曜5~6時限のクラス)の第2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://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-ss07-report02.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。

以下の問題の回答を、レポートとして A4 のレポート用紙にまとめて 5月29日の講義の初めに提出してください (レポートには結果だけを書くのではなく、問題自身や、全体的な説明、計算の途中経過の説明なども、よく分るよう工夫して、できるだけ詳しく書いてください)。なおレポートは返却しませんので自分用のコピーをとっておいてください。

なお、レポートの提出期限後に、模範解答 (解答例) を：

<http://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-ss07-report02.pdf>

に置く予定です。レポートの模範解答以外にも、講義に関連した教材や、教科書の正誤表などが：

<http://math.cs.kitami-it.ac.jp/~fuchino/chubu/index.html>

からリンクをたどってダウンロードできるようにしますのでチェックしてください。

-
1. 教科書 p.34 の体力測定のデータで、ボール投げの飛距離と 50m 走のタイムの間の相関係数を求めてください。
 2. 教科書 p.45 の演習問題 1.3.1 でのデータの相関図を作成してください。このデータでの y の x への回帰直線を求めて、相関図に書き込んでください。
 3. さいころ D_1, D_2 を 2 つ投げたとき、 D_1 と D_2 の出た目をそれぞれ n_1, n_2 として、 X を、 n_1 が偶数なら n_2 を返し、 n_1 が奇数なら、 $n_1 + n_2$ を返すような確率変数とします。このとき、
 - (a) $P(X = 1), P(X = 2), \dots$, は何になるかを考察して、確率分布表にまとめてください。
 - (b) $P(2 \leq X \leq 8)$ 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。
 - (f) (おまけの問題) 2 つのさいころを D_1, D_2 と D というアルファベットの名前にしたのはなぜだと思いますか?

1. 教科書 p.34 の体力測定で、ボール投げの飛距離と 50m 走のタイムの間の相関係数を求めてください。

k 番目の個体のボール投げの飛距離と 50m 走のタイムをあらわす変量をそれぞれ x_k, y_k ($1 \leq k \leq 10$) とすると、データは、

x_k	16	11	14	16	18	16	18	21	14	19
y_k	8.5	8.4	8.9	9.1	8.0	9.8	8.6	7.9	8.5	9.4

で、相関係数は

$$r(x, y) = \frac{N \sum_{k=1}^N x_k y_k - \left(\sum_{k=1}^N x_k \right) \left(\sum_{k=1}^N y_k \right)}{\sqrt{\left(N \sum_{k=1}^N (x_k)^2 - \left(\sum_{k=1}^N x_k \right)^2 \right) \left(N \sum_{k=1}^N (y_k)^2 - \left(\sum_{k=1}^N y_k \right)^2 \right)}}$$

で計算できる (教科書 pp.38 ~ 39) . ただし、ここでのデータのサイズ N は 10 である .

$$\sum_{k=1}^N x_k y_k = 16 \times 8.5 + 11 \times 8.4 + 14 \times 8.9 \cdots = 1417.7$$

$$\sum_{k=1}^N x_k = 16 + 11 + 14 + \cdots = 163, \quad \sum_{k=1}^N y_k = 8.5 + 8.4 + 8.9 + \cdots = 87.1$$

$$\sum_{k=1}^N (x_k)^2 = 16^2 + 11^2 + 14^2 + \cdots = 2731, \quad \sum_{k=1}^N (y_k)^2 = 8.5^2 + 8.4^2 + 8.9^2 + \cdots = 761.85$$

だから、

$$r(x, y) = \frac{10 \times 1417.7 - 163 \times 87.1}{\sqrt{(10 \times 2731 - 163^2)(10 \times 761.85 - 87.1^2)}} \approx -0.13$$

となります . この値は 0 に近いので、 x と y は相関関係があまりなさそうであると結論できます .

2. 教科書 p.45 の演習問題 1.3.1 でのデータの相関図を作成してください . このデータでの y の x への回帰直線を求めて、相関図に書き込んでください .

データの変量 x, y の一つ一つの値の組を (x_i, y_i) $i = 1, \dots, N$ (このデータでは $i = 1, \dots, 10$) , これらの平均をそれぞれ \bar{x}, \bar{y} とすると、 y の x への回帰直線は、点 (\bar{x}, \bar{y}) を通り、傾き

$$(*) \quad a = \frac{C(x, y)}{\sigma^2(x)} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{N \sum_{i=1}^N (x_i)^2 - \left(\sum_{i=1}^N x_i \right)^2}$$

を持つ直線として求められる (教科書 p.43) . データとして与えられているのは、

呼吸数 x	21	23	23	22	26	20	29	26	28	26
脈拍数 y	68	69	66	74	74	63	80	79	73	68

だから、

$$\sum_{i=1}^N x_i y_i = 21 \times 68 + 23 \times 69 + 23 \times 66 + \cdots = 17531$$

$$\sum_{i=1}^N x_i = 21 + 23 + 23 + \cdots = 244$$

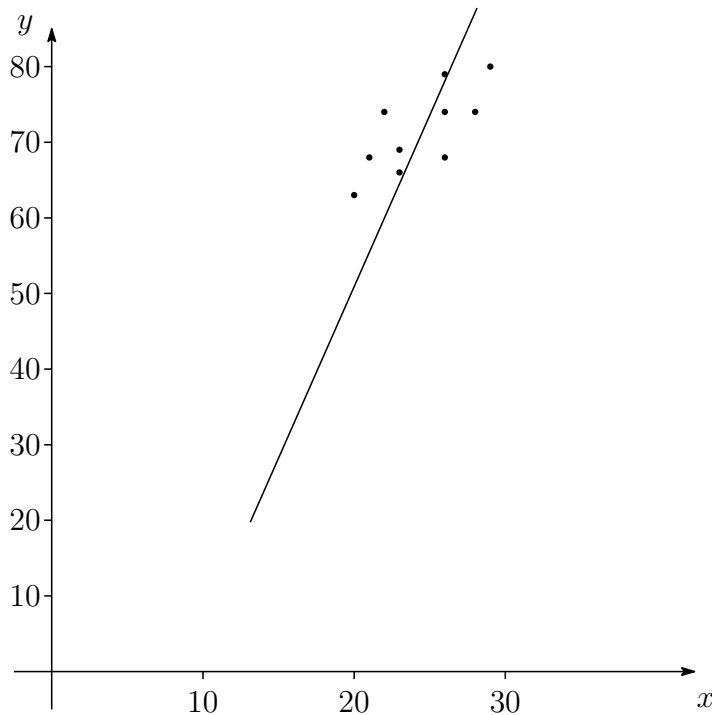
$$\sum_{i=1}^N y_i = 68 + 69 + 66 + \cdots = 714$$

$$\sum_{i=1}^N (x_i)^2 = 21^2 + 23^2 + 23^2 + \cdots = 6036$$

を (*) に代入して,

$$a = \frac{10 \times 17531 - 244 \times 714}{10 \times 6036 - 244^2} \approx 1.33$$

となる. また, $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = 24.4$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = 71.4$ である. したがって, 回帰直線は $y - 71.4 = 1.33(x - 24.4)$ で表される直線となる. この式に, たとえば $x = 0$ を代入すると, $y = 1.33 \times 24.4 - 71.4 \approx -39.0$ となるから, この直線は $(24.4, 71.4)$ と $(0, -39.0)$ を通る直線として作図できる. 以上から, y の x への回帰直線を相関図に書き込んだものは次のようになる:



3. さいころ D_1, D_2 を 2 つ投げたとき, D_1 と D_2 の出た目をそれぞれ n_1, n_2 として, X を, n_1 が偶数なら n_2 を返し, n_1 が奇数なら, $n_1 + n_2$ を返すような確率変数とします. このとき,

(a) $P(X = 1), P(X = 2), \dots$, は何になるかを考察して, 確率分布表にまとめてください.

(b) $P(2 \leq X \leq 8)$ 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.

(f) (おまけの問題) 2 つのさいころを D_1, D_2 と D というアルファベットの名前にしたのはなぜだと思いますか?

(a): まず教科書 p.48 表と同様の X の値の表を作成してみる .

$D_1 \setminus D_2$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	1	2	3	4	5	6
3	4	5	6	7	8	9
4	1	2	3	4	5	6
5	6	7	8	9	10	11
6	1	2	3	4	5	6

以上から X の確率分布表は , 次のようになることがわかる:

x	1	2	3	4	5	6	7	8	9	10	11
$P(X = x)$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

(b):

$$\begin{aligned}
 P(2 \leq D \leq 8) &= P(D=2) + P(D=3) + P(D=4) + P(D=5) + P(D=6) + P(D=7) + P(D=8) \\
 &= \frac{4}{36} + \frac{4}{36} + \frac{5}{36} + \frac{5}{36} + \frac{6}{36} + \frac{3}{36} + \frac{2}{36} = \frac{29}{36}
 \end{aligned}$$

(c):

$$\begin{aligned}
 E(X) &= 1 \times \frac{3}{36} + 2 \times \frac{4}{36} + 3 \times \frac{4}{36} + \dots = 5 \\
 V(X) &= (1-5)^2 \times \frac{3}{36} + (2-5)^2 \times \frac{4}{36} + (3-5)^2 \times \frac{4}{36} + \dots = \frac{13}{2} = 6.5 \\
 \sigma(X) &= \sqrt{V(X)} = \sqrt{\frac{13}{2}} \approx 2.55
 \end{aligned}$$

(d): 教科書 p.55 の「ポイント」の (3) を使うと , $V(3X+2) = 3^2V(X) = 9 \times \frac{13}{2} = \frac{117}{2} = 58.5$,
 $\sigma(3X+2) = \sqrt{\frac{117}{2}} \approx 7.65$

(e): (a) で求めた確率分布表を拡張して

x	1	2	3	4	5	6	7	8	9	10	11
$y = 2x^2 - x + 3$	4	9	18	31	48	69	94	123	156	193	234
$P(X = x) = P(2X^2 - X + 3 = y)$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

とできるから ,

$$E(2X^2 - X + 3) = 4 \times \frac{3}{36} + 9 \times \frac{4}{36} + 18 \times \frac{4}{36} + \dots = 61$$

(f): 次の 3 つのタイプの解答が考えられる .

- (1) [くだらない問題なので無視する .]
- (2) [真面目に答える] 賽子 (さいころ) をあらわす , 英語の dice の頭文字である .
- (3) [おちよくる] detarame の頭文字である .

あまり勧められないが ,

- (4) [くだらない問題を出したことにプロテストする]

というのももちろん可能である .