統計の手法レポート No.1
解説と解答例

渕野 昌 (Sakaé Fuchino)

fuchino@isc.chubu.ac.jp

(2007 年 06 月 22 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://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-ss07-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．

統計の手法レポート No.1 **担当： 渕野 昌** (2007年5月2日(水))

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16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-report01.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. 右下の度数分布表のデータを用いて次の (a)〜(c) に答えなさい．(d) はチャレンジ問題です（つまり，どうしてもできなければ未回答でいいですが，みなさんが挑戦してみてくれることを期待しています）．

(a) ヒストグラムを作成してください．

(b) このデータの平均，分散，標準偏差，変動係数を求めてください．

(c) 10.75 から 13.75 の間の値をとる個体の，全体に対する割合を，求めてください．

(d) 12.5 から 13.8 の間の値をとる個体の全体に対する割合を，求めてください．ただし，各階級の中でデータの数値は均等に分布するものと仮定します．

階級	度数
10.75〜11.75	5
11.75〜12.75	8
12.75〜13.75	12
13.75〜14.75	10
14.75〜15.75	9
15.75〜16.75	3
16.75〜17.75	4

2. 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× 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．この値がチェビチェフの定理（教科書の p.30）と矛盾しないことを確認してください．

(c) この試験を実施した先生は，それぞれの素点 x を $30+\frac{7}{10}x$ 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）．

1. 右下の度数分布表のデータを用いて次の (a)〜(c) に答えなさい．(d) はチャレンジ問題です（つまり，どうしてもできなければ未回答でいいですが，みなさんが挑戦してみてくれることを期待しています）.

(a) ヒストグラムを作成してください.

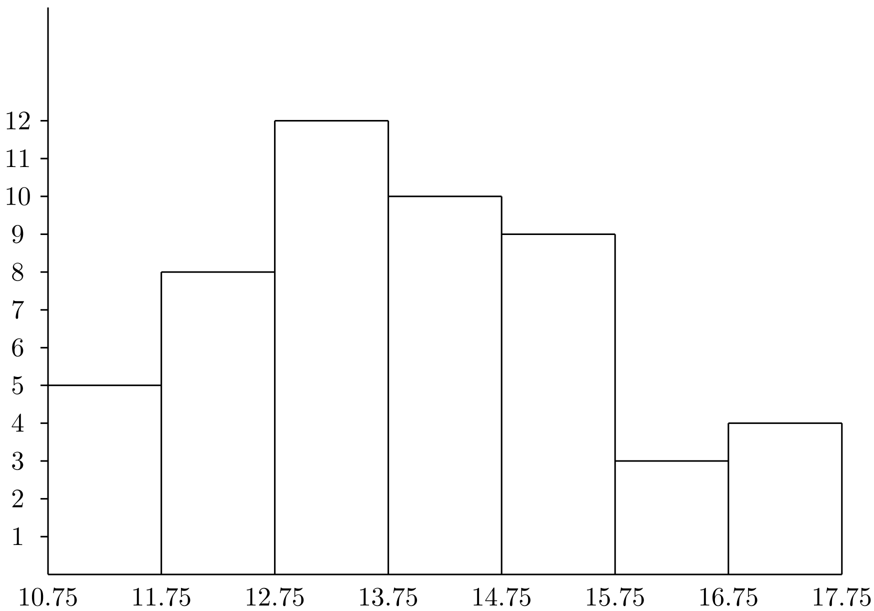
(b) このデータの平均，分散，標準偏差，変動係数を求めてください.

(c) 10.75 から 13.75 の間の値をとる個体の，全体に対する割合を，求めてください.

(d) 12.5 から 13.8 の間の値をとる個体の全体に対する割合を，求めてください．ただし，各階級の中でデータの数値は均等に分布するものと仮定します.

階級	度数
10.75〜11.75	5
11.75〜12.75	8
12.75〜13.75	12
13.75〜14.75	10
14.75〜15.75	9
15.75〜16.75	3
16.75〜17.75	4

(a):



(b): このデータのサイズは，$5+8+12+10+9+3+4=51$ である．階級 10.75〜11.75, 11.75〜12.75, ... の階級値は，それぞれ，両端の値の平均値をとると 11.25, 12.25, ... となるので，このデータの平均は,

$$\frac{1}{51}(11.25\times5+12.25\times8+13.25\times12+14.25\times10+15.25\times9+16.25\times3+17.25\times4)\approx 13.94$$

データの分散は，分散の定義を直接使って計算すると,

$$\frac{1}{51}\left((11.25-13.94)^2\times5+(12.25-13.94)^2\times8+\cdots+(17.25-13.94)^2\times4\right)\approx 2.76$$

となる.

データの各階級値の二乗の平均は,

$$\frac{1}{51}\left(11.25^2\times5+12.25^2\times8+13.25^2\times12+14.25^2\times10+15.25^2\times9+16.25^2\times3+17.25^2\times4\right)\approx 196.98$$

だから，教科書 p.28 のポイントを使うと，このデータの分散は,

$$13.94^2-196.98\approx 2.66$$

となる．1 番目の計算結果と値があまり良い合致を示していないのは，2 番目の計算で平均値を 13.94 として計算したことの誤差による．試しに 2 番目の計算をデータの平均に 2843/204 またデータの二乗の平均に 160739/816 という丸目誤差のない値を用いて行なってみると

$$160739/816-(2843/204)^2=2.764321414840452\cdots$$

となって 1 番目の計算での値とよい一致を示す.
データの標準偏差は分散の平方根だから，$\sqrt{2.76} \approx 1.66$ となる．変動係数は，この値を平均値で割って $\sqrt{2.76} \div \dfrac{2843}{204} =\approx 0.12$ となる．

(c): 12.5 から 13.75 の間の値をとる個体数は $5+8+12=25$ だから，全体に対する割合は $25/51 \approx 0.49$ である．

(d): 教科書 p.33 の中ほどで説明されている計算のやりかたによる．仮定により，12.5 から 12.75 の間の値をとる個体の数は $(12.75-12.5)/1.0\times 8$ 13.75 から 13.8 の間の値をとる個体の数は $(13.8-13.75)/1.0\times 10$ となると考えられるから，12.5 かふぁ 13.8 の間の値をとる個体の全体に対する割合は

$$((12.75-12.5)/1.0\times 8+12+(13.8-13.75)/1.0\times 10)\,/51 \approx 0.28$$

となる．

2. 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× 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．この値がチェビチェフの定理（教科書の p.30）と矛盾しないことを確認してください．

(c) この試験を実施した先生は，それぞれの素点 x を $30+\frac{7}{10}x$ 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?

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,

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

分散は,

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

標準偏差は,

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

(b): (a) から 平均点 − 2× 標準偏差 = 32, 平均点 + 2× 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 $(1-\frac{1}{2^2})50 = 37.5$ 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），$\left(\frac{7}{10}\right)^2 \times 337.6356 = 165.441444\cdots$，$\sqrt{165.441444\cdots} = 12.862404285358162\cdots \approx 13$ （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．