データの扱い方

1. 次のデータの平均, 中央値, 分散を求めよ。

(1) 3, 4, 5, 6, 7, 8, 9

(2) -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5

(3) 99.7, 101.2, 99.7, 100.4, 100.2, 99.9, 99.8, 101.0, 100.9, 100.2, 99.7

2. 次の表は, あるクラスの数学の試験の結果である。平均値と分散を求めよ。

点数	0	10	20	30	40	50	60	70	80	90	100
人数	0	0	2	4	4	19	25	26	9	1	0

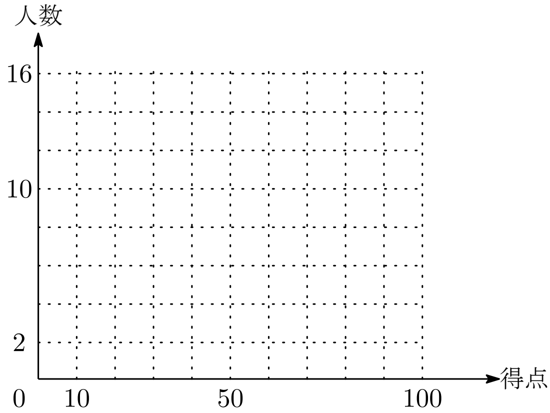
3. 次の表はある科目のテストの得点の度数分布表である。

(1) ヒストグラムを描け。

(2) 階級値を用いた近似的な平均と分散を求めよ。

(3) 中央値のある階級を答えよ。

得点	人数
10 点以上 20 点未満	1
20 点以上 30 点未満	3
30 点以上 40 点未満	7
40 点以上 50 点未満	8
50 点以上 60 点未満	15
60 点以上 70 点未満	9
70 点以上 80 点未満	4
80 点以上 90 点未満	2
90 点以上 100 点未満	1
計	50



4.　次の統計資料 A, B を全部合わせた場合の, 平均値と分散を求めよ。

統計資料	個数	平均	分散
A	20	4.5	3
B	30	7	1

$\boxed{\text{解説}}$

1.　まずは定義にしたがって求めていきます。平均を μ, 中央値を m, 分散を σ^2 と表わすことにします。

(1)

$$\begin{aligned}\mu &= \frac{1}{7}(3+4+5+6+7+8+9) = 6 \\ m &= 6 \\ \sigma^2 &= \frac{1}{7}\{(3-6)^2+(4-6)^2+\cdots+(9-6)^2\} = 4\end{aligned}$$

なお, 分布が左右対称なので, $\mu = m$ です。これで $m = 6$ としてもよいです。分散に関しては,

$$2\text{ 乗値の平均} = \frac{1}{7}(3^2+4^2+5^2+6^2+7^2+8^2+9^2) = 40$$

なので,

$$\sigma^2 = (2\text{ 乗値の平均}) - \mu^2 = 40 - 6^2 = 4$$

として求めることもできます。どちらでも求めることができるようにしておいた方がよいです。

(2)

$$\begin{aligned}\mu &= \frac{1}{11}\{(-5)+(-4)+\cdots+5\} = 0 \\ m &= 0 \\ \sigma^2 &= \frac{1}{11}\{(-5-0)^2+(-4-0)^2+\cdots+(5-0)^2\} = 10\end{aligned}$$

です。やはり (1) と同様, 分布が左右対称なので $\mu = m$, また分散に関して,

$$2\text{ 乗値の平均} = \frac{1}{11}\{(-5)^2+(-4)^2+\cdots+5^2\} = 10$$

なので,

$$\sigma^2 = (2\text{ 乗値の平均}) - \mu^2 = 10$$

としても求められます。

(3)　もちろんこのまま定義式に当てはめても解けますが, なるべく計算を簡単にするため, データを 100 からの差に変換しておきます。ついでに, 分散を簡便に求めるため, それらの 2 乗も求めています。

データ	99.7	101.2	99.7	100.4	100.2	99.9	
100 からの差	-0.3	1.2	-0.3	0.4	0.2	-0.1	
(100 からの差)^2	0.09	1.44	0.09	0.16	0.04	0.01	

	99.8	101.0	100.9	100.2	99.7	
	-0.2	1.0	0.9	0.2	-0.3	0.245455
	0.04	1.00	0.81	0.04	0.09	0.346364

右端の欄は平均です。よって表の分散は, 2 乗の平均 − 平均の 2 乗 であることを用いて,

$$2\text{ 乗の平均} - \text{平均の 2 乗} = 0.346364 - 0.245455^2 = 0.286116$$

となります。これを元データの平均と分散に戻せばよいです。平均は 100 を足せばよく, 分散はデータに一定数を足す変換をしても変化しないのでそのままです。よって, 元データの平均は 100.2455, 分散は 0.286116 となります。

次に, 中央値を求めるために, データを小さい順に並べ替えると, 次のようになります。

$$99.7,\ 99.7,\ 99.7,\ 99.8,\ 99.9,\ 100.2,\ 100.2,\ 100.4,\ 100.9,\ 101,\ 101.2$$

よって,

$$\text{中央値} = 6\text{ 番目の値} = 100.2$$

となります。

2.　今度は, データが, 一つ一つではなく, ある値が○個, 別の値が △ 個, ⋯, という形で与えられた場合です。このテストは, 15 点などの半端は無いのでしょう。

$$20,\ 20,\ 30,\ 30,\ 30,\ 30,\ 40,\ 40,\ 40,\ 40,\ \underbrace{50,\ \cdots, 50}_{19\text{ 個}},\ \cdots$$

と書き換えてもいいのですが, 一つの式として,

$$\frac{1}{90}(0\times 0 + 10\times 0 + 20\times 2 + 30\times 4 + \cdots + 90\times 1 + 100\times 0) = 60\ (\text{点})$$

としても求まります。分母の 90 は全体の人数（$= 0+0+2+4+\cdots+1+0$）です。

分散ですが,

$$\begin{aligned}\sigma^2 &= \frac{1}{90}\{(0-60)^2\times 0 + (10-60)^2\times 0 + (20-60)^2\times 2 + \cdots + (100-60)^2\times 0\}\\ &= \frac{580}{3} \approx 193\ (\text{点}^2)\end{aligned}$$

と求まります。なお, 2 乗値の平均

$$\frac{1}{90}(0^2\times 0 + 10^2\times 0 + 20^2\times 2 + 30^2\times 4 + \cdots + 90^2\times 1 + 100^2\times 0) = \frac{11380}{3}$$

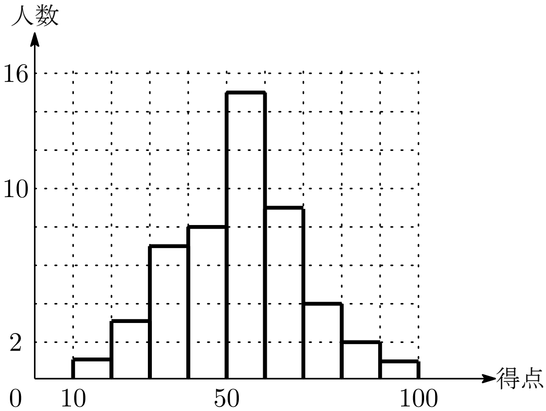
を求め,

$$\sigma^2 = 2\text{ 乗値の平均} - \text{平均の 2 乗} = \frac{580}{3}$$

としても求められます。ついでに, 最頻値は 70 点となります。

3.

(1)　ヒストグラムは以下になります。



(2)　階級値として, 各階級の中央値を用いますと, 近似的な平均は,

$$\frac{1}{50}(15 \times 1 + 25 \times 3 + \cdots + 95 \times 1) = 53.4 \text{ (点)}$$

となります。近似的な分散は, (階級値 $- 53.4)^2$ を計算するのは面倒そうなので, 2 乗値の平均

$$\frac{1}{50}(15^2 \times 1 + 25^2 \times 3 + \cdots + 95^2 \times 1) = 3125$$

を用いて,

$$2 \text{ 乗値の平均} - \text{平均の } 2 \text{ 乗} = 3125 - 53.4^2 = 273.44 \text{ (点}^2\text{)}$$

と求まります。

(3)　中央値のある階級は, テストの点が低い方から 25 番目と 26 番目の人がいる階級なので, 50 点以上 60 点未満になります。累積人数として 25 の値がある階級になります。

得点	階級値	人数	累積人数
10 点以上 20 点未満	15	1	1
20 点以上 30 点未満	25	3	4
30 点以上 40 点未満	35	7	11
40 点以上 50 点未満	45	8	19
50 点以上 60 点未満	55	15	34
60 点以上 70 点未満	65	9	43
70 点以上 80 点未満	75	4	47
80 点以上 90 点未満	85	2	49
90 点以上 100 点未満	95	1	50

4.　2 種のデータ群を統合する問題です。これも頻繁に行われることです。わからないときは次のように丁寧に考えるとよいです。

A のデータを $x_1, x_2, \cdots, x_{20}$, B のデータを $y_1, y_2, \cdots, y_{30}$ と表します。また, A の平均を $m_A(=4.5)$, 分散を $\sigma_A^2(=3)$, B の平均を $m_B(=7)$, 分散を $\sigma_B^2(=1)$ と表します。すると,

$$\mu_A = \frac{1}{20}(x_1 + \cdots + x_{20})$$
$$\mu_B = \frac{1}{30}(y_1 + \cdots + x_{30})$$

なので, 全体の平均 μ は,

$$\mu = \frac{1}{20+30}\{(x_1 + \cdots + x_{20}) + (y_1 + \cdots + y_{30})\} = \frac{1}{20+30}(20\mu_A + 30\mu_B) = 6$$

と求まります。

また, 分散に関してですが, 次のように (2 乗値の平均) − (平均の 2 乗) を用いると解きやすいです。

$$\sigma_A^2 = (A \text{ の 2 乗値の平均}) - \mu_A^2 = \frac{1}{20}(x_1^2 + \cdots + x_{20}^2) - \mu_A^2 \cdots (a)$$
$$\sigma_B^2 = (B \text{ の 2 乗値の平均}) - \mu_B^2 = \frac{1}{30}(y_1^2 + \cdots + y_{30}^2) - \mu_B^2 \cdots (b)$$

なので, 全体の分散 σ^2 を求める次の式と見比べます。

$$\sigma^2 = \frac{1}{20+30}\{(x_1^2 + \cdots + x_{20}^2) + (y_1^2 + \cdots + y_{30}^2)\} - \mu^2$$

(a), (b) より

$$(x_1^2 + \cdots + x_{20}^2) = 20(\sigma_A^2 + \mu_A^2) = 465$$
$$(y_1^2 + \cdots + y_{30}^2) = 30(\sigma_B^2 + \mu_B^2) = 1500$$

なので, σ^2 の式に代入して,

$$\sigma^2 = \frac{1}{50}(465 + 1500) - 6^2 = 3.3$$

と求まります。