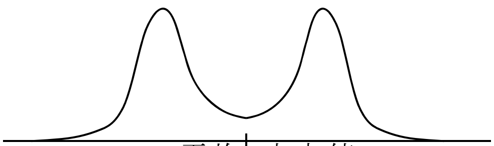
第 2 章の答え

1. 総和は 10140 から、平均は 10140/10=1014 です。最頻値 1050、中央値 (1000+1020)/2=1010 です。偏差の 2 乗和は 143240 から、それを 9(=n-1)で割ると標本分散 15915.56 となり、標本標準偏差は 126.16 です。
2. 総和Σx は、サンプルサイズ n×平均 $\bar{x}$ です。この場合、合計点は 5×70 点 =350 点です。
3. 10 人の平均は 172cm ですから、身長の総和は 172×10 です。総和から 160cm を除くと 172×10-160 です。9 人の平均は $\bar{x}$=(172×10-160)/9=173.33 です。
4. 総和は表が出た回数、平均は表が出た割合と解釈できます。
5. 加重平均は(1/3)×40+(2/3)×90=73.33 です。
6. 以下の図をみてください。明らかに左右対称の分布をしており、平均と中央値はちょうど真ん中となります。しかし、最も頻度の高いポイントである最頻値は 2 つあります。



7. ばらつきが全くない分布ですから、データは全て同じ値をとっています。
8. 1) 全ての点数に 15 点ずつ足されるため、平均は 15 点増加して 30+15=45 となる。しかし、分布は全体的に 15 点だけ右にシフトするだけなので、標本標準偏差は変わらない。2) 全ての点数を 2 倍すると、平均も 2 倍、ばらつきも 2 倍となる。平均は 30×2=60、標本標準偏差は 5×2=10 となる。3) さらに 10 点追加すると、平均は 10 点増加するため 30×2+10=70 となる。しかし、分布は全体的に 10 点だけ右にシフトするので、標本標準偏差は 5×2=10 のままである。
9. 平均気温 95°F、標本標準偏差 18°F から、摂氏では、平均と標本標準偏差はそれぞれ、

$$\bar{y} = -\frac{160}{9} + \left(\frac{5}{9}\right)95 = 35\text{、}\quad s_y = \left(\frac{5}{9}\right)18 = 10$$

となる。摂氏と華氏の定義は、教科書 p43 を参照してください。

10. 三郎君の英語の偏差値は 50+10{(75-70)/5}=60、数学の偏差値は 50+10{(66-50)/8}=70 です。よって、数学の方が良い成績といえます。偏差値の定義は、p46 を参照してください。

11. ここで a と b は $60 = a + b50$ 、 $20 = b10$ を満たすように選ばれます。よって、b=2、 a=-40 となります。たとえば、日本史の点数が 60 点なら、新しい日本史の点数は-40+2×60=80 点です。

12. a)
$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i^2 + \overline{x}^2 - 2\overline{x}x_i) = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\overline{x}^2 - 2\overline{x}\sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i^2 + n\overline{x}^2 - 2\overline{x}n\overline{x} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

b)
$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}((x_i - \overline{x}) + (\overline{x} - a))^2$$

$$= \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - a)^2 + 2(\overline{x} - a)\sum_{i=1}^{n}(x_i - \overline{x})$$

偏差の和は 0 なので、右辺第 3 項は 0 となります。偏差の和が 0 なのは、以下のように証明できます。

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i - n\overline{x} = 0$$

以上から、aに依存しているのは第 2 項だけであり、aが $\bar{x}$ のとき $\sum_{i=1}^{n}(x_i - a)^2$は最小となります[1][2]。

---

[1] この問題は、そのまま式を展開しても解くことができます。

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i^2 - 2ax_i + a^2) = \sum_{i=1}^{n} x_i^2 - 2a\sum_{i=1}^{n} x_i + na^2 = \sum_{i=1}^{n} x_i^2 - 2an\bar{x} + na^2 = \sum_{i=1}^{n} x_i^2 - \mathrm{n}\bar{x}^2 + \mathrm{n}(a - \bar{x})^2$$

ここで 3 項目だけ a に依存しており、3 項目は $a=\bar{x}$のときに値は最小となります。また、別の解法として、$\sum_{i=1}^{n}(x_i - a)^2$を aで微分して 0 と置いた式を解くことで解を求めることもできます。実際、aで微分すると$-2a\sum_{i=1}^{n}(x_i - a) = 0$となり、両辺を-2 aで割って展開すると、$\sum_{i=1}^{n} x_i - na = 0$となります。これを$a$について解くと $a=\bar{x}$となります。

[2] 問 12(b)に何の意味があるのか不思議に思う読者もおられるでしょう。議論の先取りになりますが、10 章において、データ Y と X の関係を表すモデルとして $Y_i=a+bX_i$ を考え、残差 2 乗和 $\sum_{i=1}^{n}(Y_i - a - bX_i)^2$を最小にするように未知のパラメータ a と b を推定します。この問は、もしモデルが $Y_i=a$ であれば、残差 2 乗和は$\sum_{i=1}^{n}(Y_i - a)^2$となり、これを最小にする$a$は Y の平均であることを意味しています。