STATISTICS

Form 5

Vol 11

Part 5B – Data Change

1. A	2. B	3. A	4. A	5. C	6. A	7. B

1. Note that $\{x_1, x_2, x_3, \dots, x_{20}, a_1\}$ can be formed by inserting a_1 to $\{x_1, x_2, x_3, \dots, x_{20}\}$.
   Since a_1 is the mean of $\{x_1, x_2, x_3, \dots, x_{20}\}$, we have $a_2 = a_1$ and $c_2 \le c_1$.
   However, the statement "$b_1 = b_2$" cannot be determined from the information provided.
   Thus, I only.

2. Let m, r and v be the mean, the range and the variance of the group of numbers $\{x_1, x_2, x_3, \dots, x_{80}, m_1\}$.
   Note that $\{x_1, x_2, x_3, \dots, x_{80}, m_1\}$ can be formed by inserting m_1 to $\{x_1, x_2, x_3, \dots, x_{80}\}$.
   So, we have $m = m_1$, $r = r_1$, $v \le v_1$.

   Further note that $\{x_1 + 2, x_2 + 2, x_3 + 2, \dots, x_{80} + 2, m_1 + 2\}$ can be formed by adding 2 to $\{x_1, x_2, x_3, \dots, x_{80}, m_1\}$.
   So, we have $m_2 = m + 2$, $r_2 = r$ and $v_2 = v$.
   Hence, $m_2 > m_1$, $r_2 = r_1$ and $v_2 \le v_1$.

   Thus, II only.

3. The mean of the four numbers a, a, b and b
   $= \dfrac{a+a+b+b}{4}$
   $= \dfrac{a+b}{2}$

   The required variance
   $= \dfrac{(a-\frac{a+b}{2})^2 + (a-\frac{a+b}{2})^2 + (b-\frac{a+b}{2})^2 + (b-\frac{a+b}{2})^2}{4}$
   $= \dfrac{(a-\frac{a+b}{2})^2 + (b-\frac{a+b}{2})^2}{2}$
   $= 25$

4. The standard deviation of the distribution of the scores of the class before checking the paper

$= \frac{60-50}{2}$

$= 5$

The mean of the distribution of the scores of the class after checking the paper

$= \frac{(50)(20)-60+80}{20}$

$= 51$

The standard deviation of the distribution of the scores of the class after checking the paper

$= \sqrt{\frac{(5)^2(20)-(2)[(50)(20)-60]+(2)(50)(19)+(1)^2(19)-(60-50)^2+(80-51)^2}{20}}$

$= 8$

The required standard score

$= \frac{80-51}{8}$

$= 3.625$

5. Note that the mean of the set of data $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ is m.

So, the mean remains unchanged after removing the datum m.

Since $x_4 < m < x_5$, the range remains unchanged after removing the datum m.

After removing the datum m, the sum of squares of deviation is not changed, but the number of data decreases.

So, the standard deviation increases.

Thus, I and III only.

6. Note that $\{a_1, a_2, a_3, \dots, a_{50}\}$ can be formed by inserting a_{50} to $\{a_1, a_2, a_3, a_4, \dots, a_{49}\}$.

Since $x_2 = a_{50}$, we have $x_1 = x_2$ and $z_1 \le z_2$.

However, the statement "$y_1 > y_2$" cannot be determined from the information provided.

Thus, I only.

7. Let $\overline{x}$ and v be the mean and the variance of the group of numbers $\{\alpha, \beta, \mu, \gamma, \delta\}$.
Note that $\{\alpha, \beta, \mu, \gamma, \delta\}$ can be formed by inserting μ to $\{\alpha, \beta, \gamma, \delta\}$.
Since μ is the mean of $\{\alpha, \beta, \gamma, \delta\}$, we have $\overline{x} = \mu$.

$$v = \frac{(\alpha-\mu)^2 + (\beta-\mu)^2 + (\mu-\mu)^2 + (\delta-\mu)^2 + (\gamma-\mu)^2}{5}$$

$$v = \frac{4\sigma^2}{5}$$

Further note that $\{2 - 5\alpha, 2 - 5\beta, 2 - 5\mu, 2 - 5\gamma, 2 - 5\delta\}$ can be formed by multiplying –5 to each number in $\{\alpha, \beta, \mu, \gamma, \delta\}$ and followed by adding 2 to each resulting number.

Thus, the required mean $= 2 - 5\overline{x} = 2 - 5\mu$ and the required variance $= (5)^2 v = (5)^2\left(\frac{4\sigma^2}{5}\right) = 20\sigma^2$.

8. The new mean

$$= \frac{(9)(10) - 7 - 11}{8}$$

$$= 9$$

Let $\{x_1, x_2, x_3, \ldots, x_8, 7, 11\}$ be the group of the 10 numbers.

Then, we have $\sqrt{\frac{(x_1-9)^2 + (x_2-9)^2 + (x_3-9)^2 + \cdots + (x_8-9)^2 + (7-9)^2 + (11-9)^2}{10}} = 4$.

Hence, we have $(x_1 - 9)^2 + (x_2 - 9)^2 + (x_3 - 9)^2 + \cdots + (x_8 - 9)^2 = 152$.

The required standard deviation

$$= \sqrt{\frac{(x_1-9)^2 + (x_2-9)^2 + (x_3-9)^2 + \cdots + (x_8-9)^2}{8}}$$

$$= \sqrt{\frac{152}{8}}$$

$$= \sqrt{19}$$ | r.t. 4.36

9. The mean of $\{x_1, x_2, x_3, \dots, x_8, \mu+3, \mu+1, \mu-1, \mu-3\}$

$= \dfrac{8\mu+\mu+3+\mu+1+\mu-1+\mu-3}{12}$

$= \mu$

Since the standard deviation of $\{x_1, x_2, x_3, \dots, x_8\}$ is 3, we have

$$\sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\cdots+(x_8-\mu)^2}{8}} = 3$$

Hence, we have $(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\cdots+(x_8-\mu)^2 = 72$.

The required standard deviation

$$= \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_8-\mu)^2+(\mu+3-\mu)^2+(\mu+1-\mu)^2+(\mu-1-\mu)^2+(\mu-3-\mu)^2}{12}}$$

$$= \sqrt{\frac{72+9+1+1+9}{12}}$$

$$= \sqrt{\frac{23}{3}}$$

≈ 2.77

10. (a) The lowest score in the Mathematics examination

= 92 – 54

= 38 marks

The mean score in the Mathematics examination

= 38 + 2(12)

= 62 marks

The mean score in the Additional Mathematics examination

= 88 – 2.5(10)

= 63 marks

(b) (i) Note that the score of David in the Mathematics examination is equal to the mean score. Therefore, the mean score of the class remains unchanged.

The sum of squares of deviation of the Mathematics examination is not changed, but the number of students decreases.

Therefore, the standard deviation of the scores of the class increases.

(ii) The lowest score in the Additional Mathematics examination

$= 88 - 45$

$= 43$ marks

So, David gets the lowest score in the Additional Mathematics examination.

Note that the scores of the class in the examination are distinct.

The range due to the deletion of the score of David

$< 88 - 43$

$= 45$ marks,

which means the range of the distribution of the scores of the class in the Additional Mathematics examination decreases.

Thus, the distribution of the scores of the class in the Additional Mathematics examination is not more dispersed due to the deletion of the score of David.

The claim is disagreed.

11. (a) Note that the inter-quartile range and the range of the distribution are $a - 7$ and $a - 4$ respectively.

So, we have $a - 7 > \frac{1}{3}(a - 4)$.

Hence, we have $a > 8.5$.

Note that the mode of the distribution is 9.

$$\frac{(4)(7)+(7)(12)+(9)(14)+(10)(4)+(a)(13)}{50} < 9$$

$$a < \frac{172}{13} \approx 13.23076923$$

Thus, we have $8.5 < a < \frac{172}{13}$.

Therefore, the possible values of a are 11 and 13.

(b) The required probability

$= \frac{C_1^{25} C_1^{25}}{C_2^{50}}$

$= \frac{25}{49}$

(c) After n data are added, the median of the distribution becomes 10.

So, we have $7 + 12 + 14 < 4 + 13 + n$ and $7 + 12 + 14 + 4 > 13 + n$.

Hence, we have $n > 16$ and $n < 24$.

Thus, we have $16 < n < 24$. | or $17 \leq n \leq 23$

12. Let $\{x_1, x_2, x_3, \dots, x_{10}\}$ be the group of the 10 numbers.

Then, we have $\sqrt{\dfrac{(x_1-4)^2+(x_2-4)^2+(x_3-4)^2+\cdots+(x_{10}-4)^2}{10}}=1$.

Hence, we have $(x_1-4)^2+(x_2-4)^2+(x_3-4)^2+\cdots+(x_{10}-4)^2=10$.

The mean of $\{x_1, x_2, x_3, \dots, x_{10}, 5, 7\}$

$$=\frac{(4)(10)+5+7}{12}$$

$$=\frac{13}{3}$$

The required standard deviation

$$=\sqrt{\frac{(x_1-\frac{13}{3})^2+(x_2-\frac{13}{3})^2+(x_3-\frac{13}{3})^2+\cdots+(x_{10}-\frac{13}{3})^2+(5-\frac{13}{3})^2+(7-\frac{13}{3})^2}{12}}$$

$$=\sqrt{\frac{(x_1-4-\frac{1}{3})^2+(x_2-4-\frac{1}{3})^2+(x_3-4-\frac{1}{3})^2+\cdots+(x_{10}-4-\frac{1}{3})^2+\frac{68}{9}}{12}}$$

$$=\sqrt{\frac{(x_1-4)^2+(x_2-4)^2+\cdots+(x_{10}-4)^2-(2)(\frac{1}{3})(x_1+x_2+\cdots+x_{10})-2(-4)(\frac{1}{3})(10)+(\frac{1}{3})^2(10)+\frac{68}{9}}{12}}$$

$$=\sqrt{\frac{(10)-(\frac{2}{3})(4)(10)+\frac{106}{3}}{12}}$$

$$=\sqrt{\frac{14}{9}}$$

$$=\frac{\sqrt{14}}{3}$$