(All the questions are highly unlikely to appear in the exam. One may consider skipping them if one has limited time.)

1. Fill in the blanks with the correct numbers.

(1) For the data set $\{1, 2, 3, 4, 2, 3, 4, 3, 4, 4\}$, the mean is $\square$, the mode is $\square$, and the median is $\square$.

(2) A poll carried out on a sample of 10^4 citizens obtained by random sampling among 8×10^6 citizens shows 4×10^3 are in the affirmative. The total number of citizens in the affirmative for the whole city is estimated as $\square$.

(3) The length of tongue obeys the normal distribution. The average length of tongue is known as 8 cm, with a standard deviation 1 cm. If the percentage of values lie within one standard deviation in the normal distribution is 68%, then the percentage of population with a tongue longer than 7 cm is $\square$.

(4) If all member in a data set is multiplied by 4 and the added by 2, then the mean of the data set is multiplied by $\square$ and then added by $\square$, and the variance is multiplied by $\square$ and then added by $\square$.

(5) A student scored 46 in Japanese while the average score and standard deviation are 30 and 4. The average score and standard derivation are 20 and 6 for Mathematics. If the student performed equally well in both exams, the the student scored $\square$ in Mathematics.

Q2 The table below shows the scores of 10 students for Question A and B in the examination, where some entries are missed.

Question A	32	52	62	42	(1-1)	12	72	72	82	32
Question B	32	52	62	42	52	12	72	72	82	(1-2)

By using the data, answer the following questions and fill in the blanks with the answers to the following questions.

(1) If the two sets of data have the same mean and standard deviation, fill in the missing entries to the nearest integer.

(2) Find the mean and the standard deviation of the data sets.

(3) Find the correlation coefficient between two data sets.

(1) (1-1)[ ] (1-2)[ ]

(2) Mean=[ ] Standard deviation=[ ]

(3) [ ]

Q3 The height of trees in a certain region varies from 0 m to 12 m with the probability density function $f(x) = kx$ $(0 \leq x \leq 12)$. Fill in the blanks with the answers to the following questions.

(1) Find the constant k.

(2) Find the probability of a tree in the region having a height between 4 m to 5 m.

(3) If there are 200 tress having a height between 4 m to 5 m, estimate the total number of trees in the region.

(1) $k=$ ☐

(2) ☐

(3) ☐

Brief Solutions

Q1(1) Ref: Not specific

Question related to data analysis.

Mean=$\frac{1+2\cdot 2+3\cdot 3+4\cdot 4}{10} = \boxed{3}$.

Mode=$\boxed{4}$

Rearrange the data: $\{1, 2, 2, 3, 3, 3, 4, 4, 4, 4\}$, median=$\boxed{3}$.

Q1(2) Ref: Not specific

Question related to statistics.

$8 \times 10^6 \cdot \frac{4 \times 10^3}{10^4} = \boxed{3.2 \times 10^5}$.

Q1(3) Ref: Not specific

Question related to statistics.

Note that $7 = 8 - 1 = \mu - \sigma$.

Therefore, $\Pr\{7 < x\} = \frac{1}{2}\Pr\{\mu - \sigma < x < \mu + \sigma\} + 0.5 = 0.34 + 0.5 = 0.84$, i.e. $\boxed{84\%}$.

Q1(4) Ref: Not specific

Question related to data analysis.

$\bar{x}' = \sum \frac{x'}{N} = \sum \frac{4x+2}{N} = \boxed{4}\bar{x} + \boxed{2}$.

Similarly, $\bar{x^2}' = 16\bar{x^2} + 8\bar{x} + 4$.

$\mathrm{Var}(X)' = \bar{x^2}' - (\bar{x}')^2 = 16(\bar{x^2}' - \bar{x}') = \boxed{16}\mathrm{Var}(X) + \boxed{0}$.

Q1(5) Ref: Not specific

Question related to statistics.

Consider the standard score, $\frac{46-30}{4} = \frac{x-20}{6}$, i.e. $x = \boxed{44}$.

Q2 Ref: 2016 Math A Q3

Question related to data analysis.

(1) We introduce the data sets:

$$(3, 5, 6, 4, x, 1, 7, 7, 8, 3)$$

$$(3, 5, 6, 4, 5, 1, 7, 7, 8, y)$$

Then, the original data sets are obtained by $10x + 2$.

As the data sets have the same mean, we have $x + 3 = 5 + y$, i.e. $x - y = 2$.

Also, as they have the same standard deviation, i.e. the same variance, we have

$(x^2 - x) + (3^2 - 3) = (5^2 - 5) + (y^2 - y)$, i.e. $(x - y)(x + y - 1) = 14$.

Solving, we have $(x, y) = (\boxed{5}, \boxed{3})$.

(2) The mean of the data sets above=4.9. Therefore, the original mean=$4.9 \cdot 10 + 2 = \boxed{51}$.

The variance=$\frac{283}{10} - 4.9^2 = 4.29$, i.e. standard derivation=$\sqrt{4.29} \approx 2.1$.

Therefore, the original standard deviation=$10 \cdot 2.1 = \boxed{21}$.

(3) As the two data sets have exactly the same value, we have $r = \boxed{1}$.

Q3 Ref: Not specific

Question related to statistics.

(1) As $f(x)$ is a probability density function, the area of the region bounded by it and the x-axis is equal to 1, we have $\frac{1}{2}(12)(12k) = 1$, i.e. $k = \boxed{\frac{1}{72}}$.

(2) $\Pr\{4 < X < 5\} = \frac{1}{2}(\frac{4}{72} + \frac{5}{72})(5 - 4) = \boxed{\frac{1}{16}}$.

(3) $\frac{200}{\frac{1}{16}} = \boxed{3200}$.