Name:
Result:

1.   *(5 points)*

Tomasz is a taxi driver. He records data of his first 15 trips on a random day. The mean and standard deviation of the distances for these trips are $8.82\ km$ and $2.12\ km$ respectively.

(a) Name the sampling method used.   [1]

(b) Calculate an unbiased estimates for the mean and standard deviation of the distances of all Tomasz's trips.   [2]

(c) Given that Tomasz charges an initial 8 PLN plus 3.5 PLN per each kilometre find the unbiased estimates for the mean and standard deviation of how much Tomasz charges his clients.   [2]
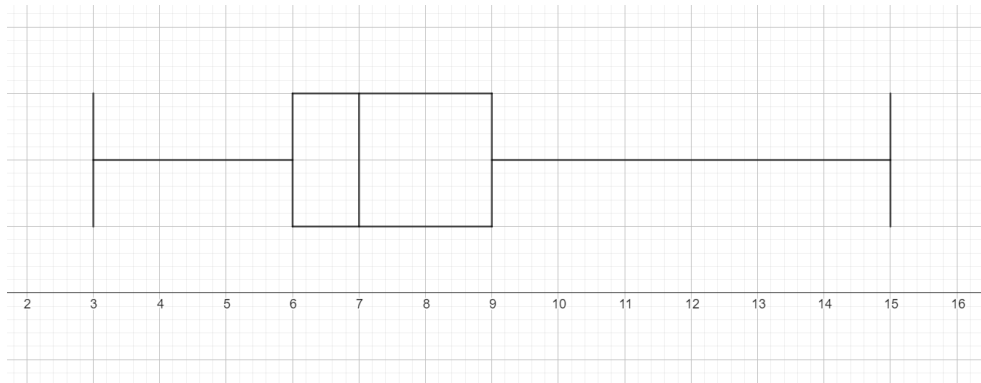
(a) Convenience sampling.

(b) $\mu = 8.82$ $s_{n-1} = \sqrt{\frac{15}{14}} \times s_n = \sqrt{\frac{15}{14}} \times 2.12 \approx 2.19$

(c) $\mu_{charge} = 3.5 \times 8.82 + 8 \approx 38.9$ $s_{n-1,charge} = 3.5 \times 2.19 \approx 7.67$

2.   *(9 points)*

A group of 120 students were asked how many hours per week they spend watching TV. The results are summarized in the diagram below:

(a) Estimate the number of students in this group who spend more than 9 hours per week watching TV.   [1]

(b) For this date find: [4]

    (i) median,

    (ii) range,

    (iii) interquartile range.

(c) Determine if the data contains any outliers. Justify your answer.   [2]

(d) Use the quartiles to estimate the standard deviation for this group.   [2]

(a) $25\% \times 120 = 30$

(b) median $= 7$, range $= 12$, IQR $= 3$

(c) outlier if smaller than $Q_1 - 1.5IQR = 6 - 4.5 = 1.5$ or greater than $Q_3 + 1.5IQR = 9 + 4.5 = 13.5$. So there are outliers as the maximum value is 15 and $15 > 13.5$.

(d) We have the following data divided into quartiles:

time [h]	frequency
$3 \leq t < 6$	30
$6 \leq t < 7$	30
$7 \leq t < 9$	30
$9 \leq t < 15$	30

Using the mid-values of the intervals we get that the standard deviation is 2.75 hours.

3. *(6 points)*

The following table shows the distribution of grades of 60 students.

Grade	Number of students	Cumulative frequency
1	1	1
2	2	3
3	2	5
4	7	12
5	**a**	30
6	21	**b**
7	9	60

(a) State the values of a and b. [2]

(b) State the modal grade. [1]

(c) Calculate the mean grade for this group. [1]

(d) A different group, consisting of 20 students, has a mean grade of 4.5. Calculate the mean grade of the two groups combined. [2]
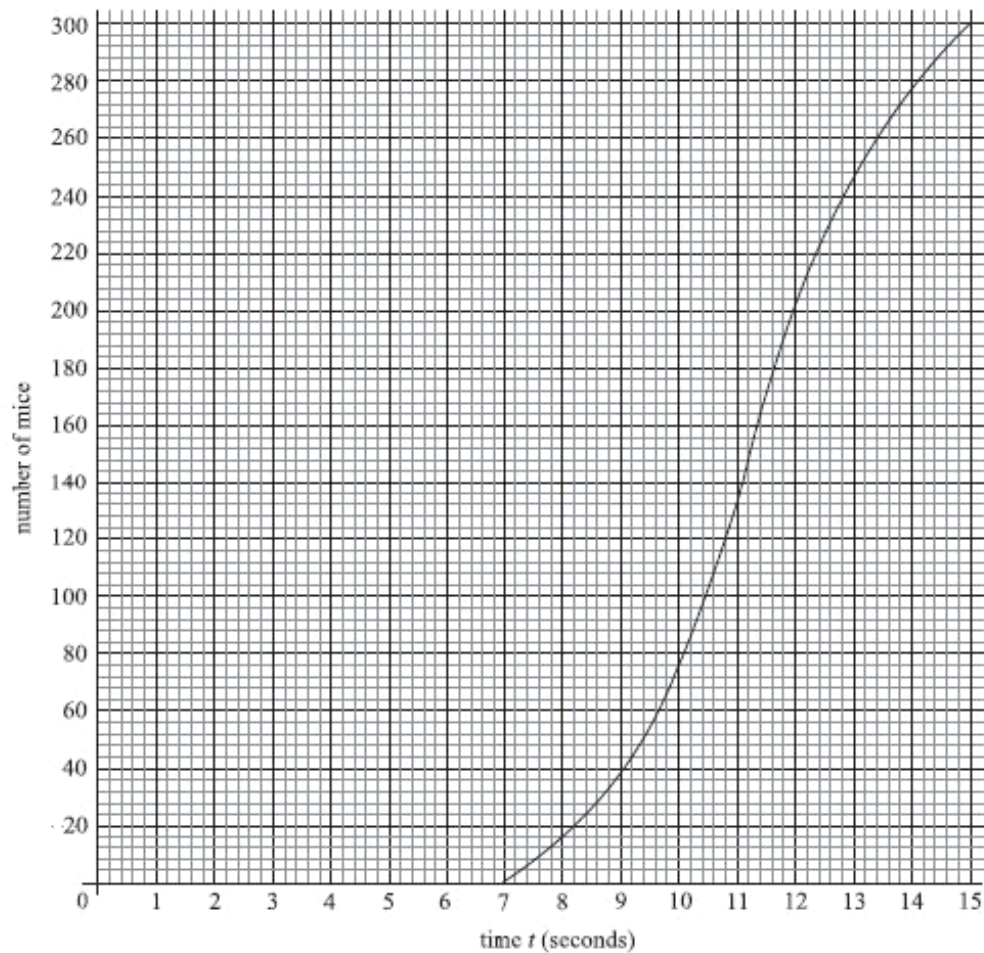
(a) $a = 18$, $b = 51$

(b) modal grade $= 6$.

(c) mean $= 5.3$

(d) mean of the two groups $= \dfrac{60 \times 5.3 + 20 \times 4.5}{80} = 5.1$

4.   *(4 points)*

In the research department of a university, 300 mice were timed as they each ran through a maze. The results are shown in the cumulative frequency diagram below.

(a) How many mice completed the maze in less than 10 seconds?   [1]

(b) The quickest 40% are given grade A, the slowest 20% are given grade C and the remaining mice are given grade B. Those that were given grade B completed the maze in between a and b seconds. Find a and b.   [3]

(a) 76

(b) $a = 10.8$ $b = 12.8$.

5.   *(6 points)*

Wanda wishes to see whether there is any correlation between a person's age and the number of objects on a tray which could be remembered after looking at them for a certain time. She obtains the following table of results.

Age (x)	25	24	45	49	31	51	42	39	65	70	54	42	29	37	28	67
Number of objects (y)	25	19	13	19	24	11	22	23	15	15	18	16	25	15	25	10

After analysing the data Wanda noticed that 10 of the participants in the study had higher education and their results differ from the remaining 6 participants.

(a) Find the regression line, in the form $y = ax + b$, only for the participants with higher education.   [2]

(b) Interpret the coefficient a in your regression line.   [1]

(c) Use your regression to estimate how many object a 40 year old with higher education will remember. Give two reasons, why your estimation is valid.   [3]

(a) The data for those with higher education looks as follows:

Age (x)	25	49	31	42	39	65	70	54	29	28
Number of objects (y)	25	19	24	22	23	15	15	18	25	25

The regression line is then $y = -0.251x + 32.0$

(b) An increase in age by 1 corresponds, on average, to a decrease in the number of remembered objects by 0.251.

(c) When $x = 40$ we get $y = 21.9 \approx 22$. Reasons: interpolation (new data within the range of the analysed data) and the correlation is very strong.