

National
Qualifications
2026

X803/77/12

**Statistics
Paper 2**

FRIDAY, 22 MAY

10:30 AM – 1:15 PM

Total marks — 90

Attempt ALL questions.

You may use a calculator.

To earn full marks you must show your working in your answers.

State the units for your answer where appropriate.

Write your answers clearly in the spaces provided in the answer booklet. The size of the space provided for an answer is not an indication of how much to write. You do not need to use all the space.

Additional space for answers is provided at the end of the answer booklet. If you use this space you must clearly identify the question number you are attempting.

Use **blue** or **black** ink.

You must leave your answer booklet on your desk; if you do not, you could lose all the marks for this paper.

You may refer to the Statistics Advanced Higher Statistical Formulae and Tables.



* X 8 0 3 7 7 1 2 *



Total marks — 90
Attempt ALL questions

1. A research paper was presented at the 2015 Australasian Piano Pedagogy Conference called 'Pianist Hand Spans: Gender and Ethnic Differences and Implications for Piano Playing' by R Boyle, et al. In this research, hand spans of 303 adult female pianists' right hands were recorded. Hand spans were measured in centimetres from the outside tip of the thumb to the outside tip of the fifth finger. A summary of the pianists' right-hand span data is given below.

	White	Asian
Sample size	216	87
Minimum	16.3	17.2
Maximum	24.1	22.9
Mean	20.2	19.8
Median	20.3	19.8
Lower quartile	19.3	19.1
Upper quartile	21.0	20.6
Interquartile range	1.7	1.5
Standard deviation	1.38	1.15

- (a) Determine whether there are any possible outliers within the 216 White pianists. 2

You may assume for this research paper that the distributions of right-hand spans are normally distributed and that the sample standard deviations are good approximations to the corresponding population standard deviations.

- (b) Perform an appropriate z -test at the 0.1% significance level on whether there is a difference in the mean hand spans between White and Asian pianists. 5
- (c) State one further assumption required for the z -test performed in part (b). 1

2. During an analysis of the attendance of 40 university students, a lecturer compared the number of lectures missed by a student over a year (x) to their mark in the final exam at the end of year (y).

The lecturer found a linear association between these two variables and the following summary statistics were obtained:

$$\sum x = 275, \sum y = 2468, S_{xx} = 1958.375, S_{xy} = -5029.5, S_{yy} = 18734.4$$

- (a) (i) Determine the equation of the least squares regression line of y on x . 2
- (ii) Find the residual for a student who missed 10 lectures and achieved a mark of 58 in the end of year exam. 2
- (b) State why it would be inappropriate to use the regression line found in part (a) to estimate the number of lectures missed by a student if their mark in the end of year exam is known. 1
3. In a charity raffle, each raffle ticket is numbered from 1 to n , where n is the number of tickets sold. One winning ticket is selected at random.
- Let X be defined as the number of the winning ticket.
- (a) State the distribution of X . 1
- (b) Given that $V(X) = 2552$, find the value of n . 2
- A local businessman agrees to donate a sum of money to the charity, in pounds sterling, that is equivalent to the winning ticket number multiplied by 20.
- Let Y be defined as the value of the donation.
- (c) Calculate the standard deviation of Y . 2

[Turn over

4. Between 2012 and 2021, the Fire Services of Scotland attended more than 265 000 fires, of which over 75 000 were for burning refuse, where 'refuse' is another word for 'rubbish'. Some of these refuse fires were started by accident, but the majority were started deliberately.

A fire safety officer thinks that the number of deliberate fires is more than 6 times the number of accidental fires.

To investigate this claim, a random sample of 300 refuse fires was taken from over the 10 year period and their day of the week was recorded, along with whether it was suspected to have started by accident, or deliberately.

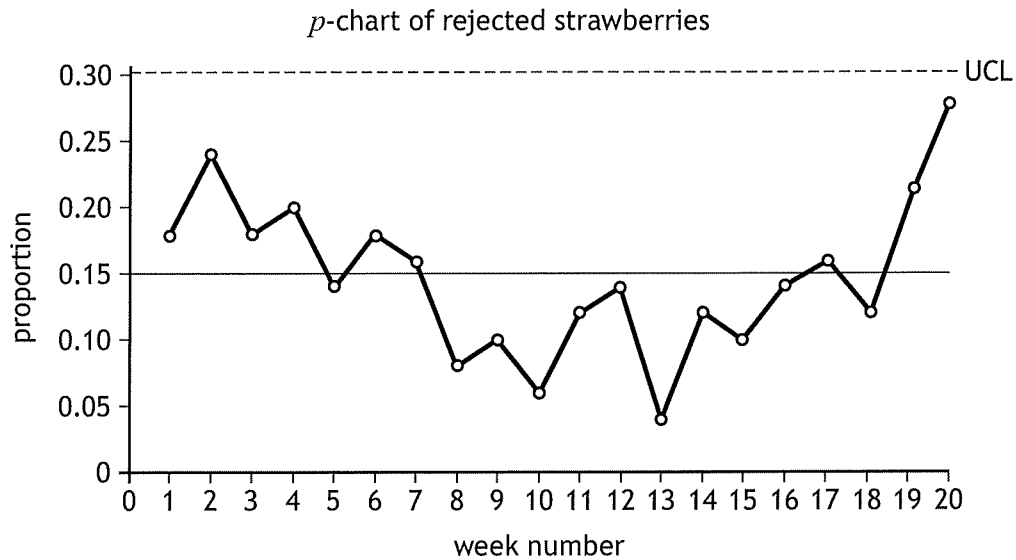
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Accidental total	5	9	7	8	4	8	9
6 × Accidental total	30	54	42	48	24	48	54
Deliberate total	38	30	32	30	36	40	44
Difference	8	-24	-10	-18	12	-8	-10

Using a 5% level of significance, conduct a Wilcoxon signed rank test to determine if, on average, the number of deliberate fires is more than 6 times the number of accidental fires.

5. A farm grows a selection of fruit and sells them to local shops. A proportion of the produce is rejected due to quality issues.

The farm expects that 15% of their strawberries will be rejected by shop A, which regularly purchases from the farm. A sample of 50 strawberries is randomly selected each week, over the 20-week growing season from May to October. The proportion deemed unsatisfactory is recorded and plotted in the p -chart below.

The 3-sigma upper control limit (UCL) is also shown.



- (a) Calculate the value of the 3-sigma limits to 4 decimal places, and hence comment on why it is not appropriate to use the calculated value of the lower control limit in this case. 3
- (b) Without further calculation, describe a feature of the chart that indicates that the process went out of statistical control, and give a reason why the farm would be pleased with this feature. 2

Shop B purchases a different variety of strawberry from the farm and past records show that it rejects 8% of them. The farm intends to use the same method of analysis as done for shop A.

- (c) Calculate the smallest sample size that the farm could take so that a normal approximation to the binomial distribution would be a valid approach. 2

[Turn over

6. An ice-cream machine should deliver portions that have a mean mass of 70 grams. The ice-cream seller suspects that the mean mass of ice-cream being served from the machine is less than 70 grams.

The masses of 15 randomly selected portions served from the ice-cream machine are recorded.

- (a) Give a reason why it would be useful to construct a graphical display of the masses of the 15 portions of ice cream.

1

The 15 portions had a mean mass of 67 grams with a standard deviation of 4 grams.

- (b) (i) State the assumption required to conduct a parametric test of whether the ice-cream seller's suspicion is justified.

1

- (ii) Assuming the assumption from part (b) (i) is satisfied, perform an appropriate parametric test using a 1% significance level.

6

7. In a small office, a kettle is used to make hot drinks during office hours. Managers are considering whether to replace the kettle with a larger water heater. To check if it is economically viable, they need to evaluate how often staff currently use the kettle per day.

It is estimated that staff boil the kettle at a mean rate of 2.5 times per hour.

- (a) Calculate the probability that the kettle is boiled exactly 8 times between the hours of 9 am and 12 noon.

2

A normal approximation is used to calculate the probability in part (a), yielding a value of 0.1425.

- (b) Comment on the use of the approximation in this context.

1

A standard working day is from 9 am to 5 pm. Managers believe that the water heater is a viable option if there is at least an 80% chance of the kettle being used more than 15 times in a day.

- (c) Use an approximation to calculate the probability that the kettle is used more than 15 times during a standard working day, and comment on whether the water heater is viable.

5

8. Researchers who were interested in the social awareness of children carried out an experiment to investigate the age at which children start to be aware of other people's preferences. A random sample of children from various backgrounds was brought into the research unit. The children were placed into two groups: a younger group of 14-month-olds and an older group of 18-month-olds.

In the experiment, a researcher would sit with a child and two bowls of food: one bowl had broccoli in it, and one bowl had crisps in it. The crisps were always preferred by the children. The researcher would pick some food from each bowl, reacting with disgust to the crisps and with enjoyment to the broccoli. The researcher would then hold out their hand and ask the child to give them one of the foods.

When the researcher held out their hand to be offered some food, 6 of the 31 younger children offered broccoli to the researcher, whilst 38 of the 55 older children offered broccoli.

Stating a required assumption, perform a z -test at the 1% significance level to assess whether there is evidence that 18-month-olds are more aware of other people's preferences than 14-month-olds.

8

9. A company produces cans of drinks which are packaged in cardboard multipacks containing 24 cans.

The mass, in grams, of an empty can is normally distributed with mean 18 g and variance 2 g^2 .

The mass, in grams, of the multipack cardboard packaging is normally distributed with mean 196 g and variance 1 g^2 .

The total mass of 24 cans filled with drink and packed in cardboard packaging is known to be normally distributed with mean of 8884 g and variance 121 g^2 .

Stating a required assumption, determine the distribution of the mass, in grams, of just the drink in a single can.

7

[Turn over

10. A consultancy firm helps match up government employees with training courses to learn new skills. Training providers who want to offer training courses need to complete an application process for the consultancy firm. This process requires the training provider to agree to some of their previous customers taking part in a survey.

The training provider has to submit the names of 20 previous customers of their own choosing to the consultancy firm, who would then be asked to take the survey. The consultancy firm uses the survey results to determine the quality of the training course and whether that training provider could then offer their course to government employees.

Describe, giving a reason, the potential problem that this method of sample selection would create.

2

11. When travelling from town G to town H using public transport there is a choice of using either a bus or train.

The bus operates 100 times a week and the probability that any randomly selected bus will arrive late to town H has been calculated to be 0.043.

The train operates 90 times a week and the probability that any randomly selected train will arrive late to town H has been calculated to be 0.09.

- (a) Assuming buses run independently, calculate the probability that at least one bus is late in a given week, stating clearly the distribution used.

3

A passenger takes the first available mode of public transport that leaves town G when they are ready to depart, and they arrive late to town H.

- (b) (i) Calculate the probability that this passenger travelled by train.
(ii) State an assumption that needed to be made in part (b) (i).

4

1

In an attempt to change the number of trains that arrive late, alterations are made to the train timetable. After a period of time, 75 journeys are randomly selected and it is noted that 6 of them recorded the train being late to town H.

- (c) (i) Calculate an approximate 95% confidence interval for the true proportion of trains that are late after the alterations to the timetable.
(ii) Comment, with reason, on the attempt to change the number of trains that arrive late.
(iii) State the two conditions that are required for calculating this interval and state why the interval is only approximate.

3

2

2

12. X has the distribution $X \sim N(\mu, \sigma^2)$.

(a) Calculate $P(\mu - 0.2\sigma < X < \mu + 0.2\sigma)$.

3

A sample of size 36 is drawn from the population X .

(b) (i) State the distribution of \bar{X} .

1

(ii) Calculate $P(\mu - 0.2\sigma < \bar{X} < \mu + 0.2\sigma)$.

3

(c) Determine an inequality that the sample size would have to satisfy, in order to give

$$P(\bar{X} < \mu + 0.2\sigma) \leq 0.9$$

4

[END OF QUESTION PAPER]

[BLANK PAGE]

DO NOT WRITE ON THIS PAGE

[BLANK PAGE]

DO NOT WRITE ON THIS PAGE

[BLANK PAGE]

DO NOT WRITE ON THIS PAGE