National Qualifications 2019

X703/77/11 Statistics

TUESDAY, 7 MAY
1:00 PM – 4:00 PM

Total marks — 100

Attempt ALL questions.

You may use a calculator.

Full credit will be given only to solutions that contain appropriate working.

State the units for your answer where appropriate.

Write your answers clearly in the answer booklet provided. In the answer booklet you must clearly identify the question number you are attempting.

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

Before leaving the examination room you must give your answer booklet to the Invigilator; if you do not, you may lose all the marks for this paper.

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

MARKS

Total marks — 100

Attempt ALL questions

1. On the main Glasgow to Edinburgh motorway, breakdowns occur at a mean rate of 3 per day. Calculate the probability that

(a) fewer than the mean number of breakdowns occur in a particular day **2**

(b) exactly 25 breakdowns occur in a given 7-day week. **2**

2. A study in which 1000 people were tested for colour blindness also took note of their handedness (left or right). The results are recorded in the contingency table below.

		Handedness	
		Left handed	**Right handed**
Colour blind	**No**	130	780
	Yes	10	80

(a) Given that a person chosen at random is colour blind, calculate the probability that they are left handed. **1**

(b) Use probability theory to determine whether being colour blind is independent of being left handed. **4**

Colour blindness is thought to be associated with gender.

The table shows the distribution of colour blindness in this study.

		Male	**Female**
Colour blind	**No**	450	506
	Yes	40	4

(c) Perform a chi-squared test of association between colour blindness and gender.

Comment on the result in the context of this study. **7**

MARKS

3. A researcher wishes to investigate the proportion of children's books that have a female central character.

To investigate this, the researcher decides to look at a sample of children's books. Permission is obtained to examine all of the 100 titles in the children's section of a local bookshop, to determine the number of female central characters.

(a) Name the sampling strategy chosen and state a disadvantage of this method, together with its possible consequence for the researcher. **3**

(b) Describe theoretically how a systematic random sample of 4% of children's books in a large city bookshop could be taken. **3**

[Turn over

MARKS

4. A football fan claims that the first team to score a goal is more likely to win the match. The fan collected data on two teams over one season.

The table summarises the games in which the team won.

	Season	
	Team A	**Team B**
Number of games in which team scored first	22	15
Number of games in which team conceded first	11	16
Total number of games in season	33	31

Use a parametric method to assess the evidence for the football fan's claim. **5**

5. A random sample of size 25 is taken from a continuous uniform distribution, defined over the interval 7 to 22.

(a) Use the Central Limit Theorem to approximate the probability that the sample mean is greater than 16·7. **3**

(b) Justify the use of the Central Limit Theorem. **1**

MARKS

6. Wood burning stoves are becoming a common addition to many homes. However, the heat output of stoves is known to a researcher to be affected by the type of wood used. The table below details the mean energy in kilowatt hours (kWh) produced from burning 1 solid cubic metre of various species of tree.

Species	Mean energy
Holly	3591
Beech	3032
Oak	2926
Ash	2926
Elm	2873
Maple	2820
Birch	2660
Pine	2341

(a) By considering the types of data in the table above, describe, with justification, a suitable way to display this information graphically. **2**

(b) The researcher selects 10 random samples of firewood from the same tree of unknown species. These were burnt and the energy produced was measured and is recorded below.

2725 2650 2421 2793 2239 3225 2156 2692 2369 2725

with $\sum x = 25995$ and $\sum x^2 = 68456947$

Stating one necessary assumption, calculate a 95% confidence interval for the population mean kWh produced. **5**

(c) Using the table above, suggest the species of tree that the 10 samples may have come from, justifying your choice. **2**

(d) The researcher wishes to be more confident in her conclusion. Calculate a 99% confidence interval and suggest how it would affect her conclusion. **3**

[Turn over

MARKS

7. A business is starting up to produce small tubes of glue. The quality assurance manager wishes to monitor the weight of glue (grams) in the tubes and decides that it will be appropriate to weigh 6 tubes, selected at random every day, in order to create a control chart for tube weights.

Results from the first 8 days are shown below.

Day	Tube weights X						Daily mean $\overline{X}$	Daily range R
1	29·0	30·0	28·0	28·6	31·0	28·0	29·1	3·0
2	30·0	30·8	29·0	30·0	31·0	31·0	30·3	2·0
3	29·0	30·0	31·0	30·0	31·0	32·0	30·5	3·0
4	29·0	30·0	28·8	29·0	30·0	32·0	29·8	3·2
5	30·0	29·5	28·0	28·0	29·0	29·5	29·0	2·0
6	29·0	30·0	28·0	28·6	30·0	29·0	29·1	2·0
7	29·0	30·0	29·9	29·0	29·0	29·5	29·4	1·0
8	30·0	29·5	28·2	29·0	29·5	29·0	29·2	1·8

As this is the first time this process has been implemented, the following information is used to calculate a target value and 3-sigma limits from the first 8 days' data, with the range being used to calculate an estimate of the standard deviation of tube weight.

The target value is denoted by $\overline{\overline{X}}$ and the 3-sigma limits by $\overline{\overline{X}} \pm 3\frac{\hat{\sigma}}{\sqrt{n}}$

where $\overline{\overline{X}}$ = mean of the daily means $\overline{X}$

$\hat{\sigma} = \frac{\overline{R}}{d}$ ($\hat{\sigma}$, estimated standard deviation of X)

$\overline{R}$ mean of the daily ranges R

and d is a scaling factor that varies with sample size n.

(a) Using $d = 2{\cdot}534$, calculate the target value and the 3-sigma limits for the chart. **3**

(b) A second glue is manufactured with tubes of similar weight. To monitor this process, the quality assurance manager decides to randomly sample 9 tubes of glue each day to provide data for an *x*-bar chart for mean tube weight.

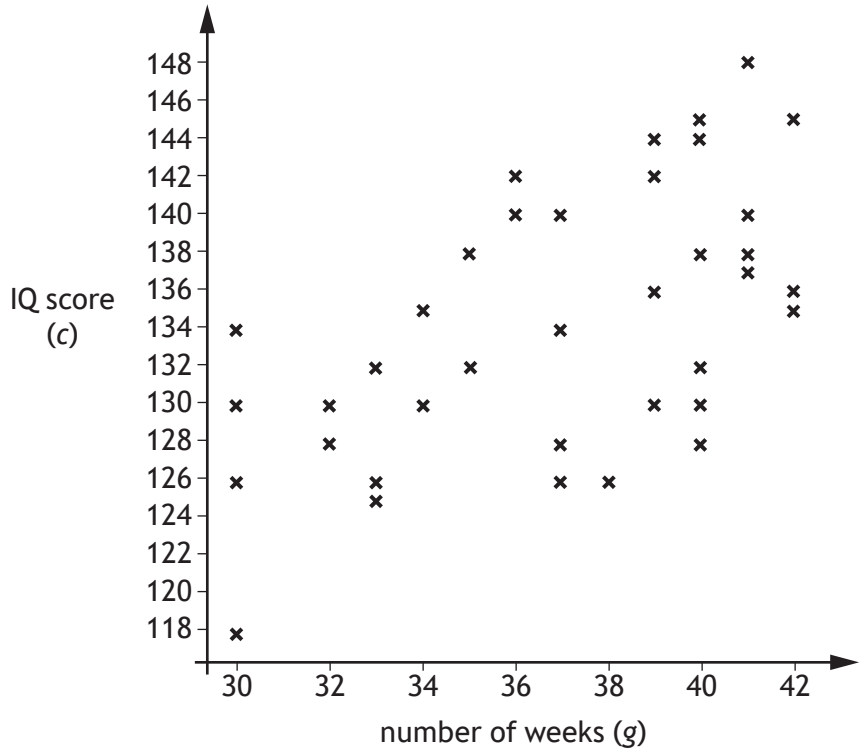
The results for the first 8 days yield $\overline{\overline{X}} = 29{\cdot}92$ and $\overline{R} = 2{\cdot}35$.

If the calculated upper 3-sigma limit for this process is 30·71, determine the value of the scaling factor, d, that has been used for this process. **3**

MARKS

8. Gestation is the period of time from conception to birth.

In a study looking at the relationship between the number of weeks of gestation (g) and a child's IQ score at age 12 years (c), the following scatterplot was created from a sample of 37 suitable children.



(a) From the scatterplot, comment on the relationship between the number of weeks gestation and the IQ score of the 12-year-olds sampled. **1**

(b) The following statistics were calculated from the data displayed above.

$S_{gg} = 531{\cdot}5676 \quad S_{gc} = 555{\cdot}0811 \quad S_{cc} = 1731{\cdot}2973$

Calculate the coefficient of determination and state what this measures with respect to a linear model. **2**

(c) Perform a suitable hypothesis test at the 1% significance level to investigate the linear association between the number of weeks of gestation and the IQ at age 12, stating any assumption required. **6**

(d) State an additional check that should be done after fitting a linear model, and what it should reveal. **2**

[Turn over

MARKS

9. In a particular year the national results of a Statistics examination had a median of 65.

The following year the assessment team initially marked a random sample of 22 candidates' scripts and awarded the following marks.

86 80 78 73 69 65 62 61 59 58 54
51 49 47 43 40 38 37 35 32 29 29

The team wished to determine if there was any evidence from this sample of poorer performance than in the previous year.

Stating a necessary assumption, perform a Wilcoxon Signed-Rank test to assess the evidence. **9**

10. The wingspan (cm) of a particular species of seabird is thought to be distributed approximately N(50,16). In order to research changes in wingspan for this species, a marine biologist has taken a random sample of 25 seabirds. The mean wingspan for this sample is found to be 48·3 cm.

(a) Perform a test at both the 5% and 1% levels of significance to assess whether there is evidence that the wingspan of this species has decreased. **5**

The marine biologist is planning to take many more random samples of 25 birds. She would like to know that if $\overline{x} \leq b$ then there is evidence that the wingspan of the species has decreased.

(b) Calculate the values of b for each of the levels of significance above. What instruction would you give the marine biologist, concerning the conclusion she should come to depending on the value of any of her sample means? **4**

In real life, a population variance is rarely known.

(c) Explain how you would proceed to test for a decrease in wingspan, as in part (a), if the mean wingspan is known to be 50 cm and the population variance is unknown.

(You are not required to carry out the test or to present a conclusion). **3**

MARKS

11. There are two available driving instructors at a local driving school and we may assume that learner drivers do not change instructors.

Drivers taught by Instructor A have a 55% chance of passing first time. Of those who do not pass first time, 65% usually go on to pass on their second attempt.

(a) Find the proportion of these drivers who will have passed after one or two attempts. **2**

Of all those taught by Instructor B, 52% typically pass first time, while 35% fail first time **and** go on to pass on their second attempt.

(b) (i) Instructor B claims his pass rate after one or two attempts is better than that of Instructor A.

Justify whether or not he is correct.

(ii) Of those who fail the first time with Instructor B, what percentage are expected to pass on their second attempt? **4**

Instructor B is found to have twice as many learner drivers as Instructor A.

(c) Calculate the probability that a learner who has failed their driving test after two attempts was taught by Instructor B. **3**

12. In a 'low stakes' area of a Las Vegas casino, a player pays 1 dollar to play a game where three unbiased regular octahedral dice with faces marked 1 to 8 are thrown.

- If all 3 dice show a 1 the player receives 100 dollars
- If 2 dice show a 1 the player receives 10 dollars
- If only 1 dice shows a 1 the player receives 1 dollar
- Otherwise the player receives nothing

The random variable X represents the player's profit for one game.

(a) Tabulate the probability distribution of X, with probabilities correct to 4 decimal places, and show that $\mathrm{E}(X) = -0{\cdot}1029$ and $\mathrm{SD}(X) = 4{\cdot}8562$. **5**

In a 'high stakes' area of the casino a different game is played, yielding $\mathrm{E}(Y) = -0{\cdot}06$ and $\mathrm{SD}(Y) = 20$, where Y represents the player's profit for one game.

A gambler plays the low stakes game 60 times and the high stakes game 45 times.

(b) (i) Calculate the expected value and standard deviation of his winnings, stating an assumption you have made.

(ii) Comment on these values within the context of the question. **5**

[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

FOR OFFICIAL USE

National Qualifications

Mark

X703/77/01

Statistics
Answer booklet

X7037701

Fill in these boxes and read what is printed below.

Full name of centre

Town

Forename(s)

Surname

Number of seat

Date of birth

Day Month Year

Scottish candidate number

Write your answers clearly in this answer booklet. You must clearly identify in the margin the question number you are attempting.

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

Before leaving the examination room you must give this booklet to the Invigilator; if you do not you may lose all the marks for this paper.

X703770101

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

X7037701102

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

For marker's use	
Question	Marks