National Qualifications 2018

X703/77/11 Statistics

THURSDAY, 10 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 which 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. Two families with newborn babies compare sleep patterns and count the number of times per night that each baby wakes up. Family A's baby wakes up with a mean rate of 3 times per night and family B's baby wakes up with a mean rate of 2 times per night.

(a) Calculate the probability that family B's baby wakes up exactly 3 times in any given night. **2**

(b) Calculate the probability that, during a given night, neither baby wakes up. **2**

2. A school management team has gathered together information on the percentages of pupils and staff going on school trips, and the percentage of pupils and staff who arrive late for these trips. For one particular trip, of those going 46% were junior pupils, 41% were senior pupils and the rest were staff. It is noted that 9% of those junior pupils, 20% of the senior pupils and 6% of the staff arrived late.

(a) Calculate the probability that a randomly chosen person on this trip:

(i) was a junior pupil who arrived on time

(ii) was late. **3**

(b) If a person arrived late, calculate the probability that it was a senior pupil. **2**

The school management team also wishes to canvass the opinions of pupils and staff on the value of school trips. Practical considerations determine that only 10% of the school population can be surveyed. It is felt that there will be a difference of opinions between staff, junior pupils and senior pupils.

(c) How would you advise the school management team to conduct the survey? **2**

3. A naturalist is interested in the changes in average density of a particular species in a local area. He has recorded the average density of the species in two different habitats over the past seven years, as shown below.

Year	1	2	3	4	5	6	7
Hab 1 (x)	313	342	366	350	376	438	400
Hab 2 (y)	83	94	91	86	102	113	98

The data can be summarised by

$$S_{xx} = 10005{\cdot}4 \quad S_{xy} = 2265{\cdot}43 \quad S_{yy} = 623{\cdot}429$$

and a scatterplot of the data shows that a linear relationship is appropriate.

The naturalist wishes to continue observations, but finds it safer to do so in the first habitat than in the second and so, in future, he wishes to use least squares linear regression to estimate the density in Hab 2 from that of Hab 1.

(a) Test the hypothesis that the slope parameter of the least squares regression line is zero. **7**

(b) Before recommending this model explain what else you would calculate in order to establish its usefulness.

Give a reason for your answer. **2**

4. A manufacturer claims the life expectancy of a car headlight bulb is normally distributed with mean 10 000 hours and standard deviation 250 hours.

(a) Find the probability that a new bulb lasts at least 10 100 hours. **2**

(b) If the company wants to ensure 90% of bulbs meet their claim, what value for life expectancy should the company print on the packaging? **3**

(c) The weight (grams) of each bulb is N(24, 1). Each bulb is in its own box, the weight of which is N(5, 0·5). One hundred bulbs are packed in a crate whose weight is N(75, 7).

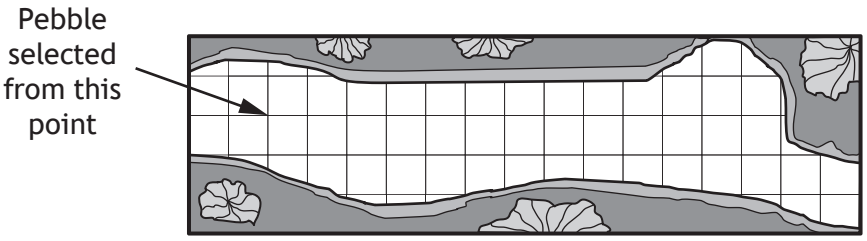
Stating any assumption required, calculate the probability that the total weight of a crate is less than 3 kg. **6**

[Turn over

MARKS

5. In 1954, a new sampling method was introduced to determine the distribution of pebble sizes in a stream bed, the size being the longest dimension (mm).

A grid system is established over the stream bed and one pebble is picked up at every intersection of grid lines that lies within the stream bed, as illustrated in the diagram below.



(a) State the type of sampling method used. **1**

Before this new method was developed it was usual to take bulk samples (many pebbles from the same location) from one or two locations in the area of the stream bed to be sampled.

(b) Give a reason why this older method might lead to the taking of unrepresentative samples. **1**

A researcher was asked to use the new method at a stream bed in Virginia in North America where it was known using the older method that pebble size had a mean of 115·3 mm with standard deviation 21·6. The researcher samples 100 pebbles and calculates a mean size of 119·4 mm.

(c) Assuming that the standard deviation of pebble size has not changed, obtain a 90% confidence interval for the mean pebble size from the new method.

Interpret this confidence interval in the context of the two sampling methods. **5**

6. A Mann-Whitney test is to be performed using two independent samples of sizes 4 and 7. The sum of ranks of the smaller sample is 14.

By listing all the subsets of appropriate ranks, show that

$$\mathrm{P}(\text{sum of ranks} \leq 14) = \frac{2}{55}.$$ **5**

MARKS

7. Tulip bulbs are sold in packets of 6 and the growers claim that only 10% of such bulbs will fail to grow. A random sample of 120 such packets yielded the following frequencies of the number of bulbs in a packet which failed to grow.

Number of failed bulbs	0	1	2	3	4	5	6
Frequency	59	38	19	3	1	0	0

(a) Perform a test to assess the evidence that the data fits a binomial distribution with $p = 0{\cdot}1$. **7**

(b) Explain why an assumption underlying the binomial model in this context of tulip bulbs might be inappropriate. **1**

8. Data obtained from STATS19 (the national database of police-reported injury road collisions in Great Britain) for the years 2008–2012 revealed that, nationally, 11·9% of accidents involved a young driver.

For a new stretch of road in northern Scotland, the local council keeps accident data. In 2013, 100 accidents were recorded at this location, with 18 involving a young driver.

(a) Perform an appropriate hypothesis test to determine if the council's data provide any evidence, at the 5% significance level, that the proportion of accidents involving young drivers at this location for 2013 is significantly greater than the 2008–2012 national figure would lead you to expect. **6**

At the end of 2013, safety measures aimed at young drivers were introduced at this location. Council records for a four-month period in 2014 showed 40 accidents at this location.

Let d be the maximum number of young drivers involved in these 40 accidents for the council to conclude, at the 5% level, that the proportion of accidents involving young drivers at this location is significantly less than the 2008–2012 national figure.

(b) Determine the value of d. **3**

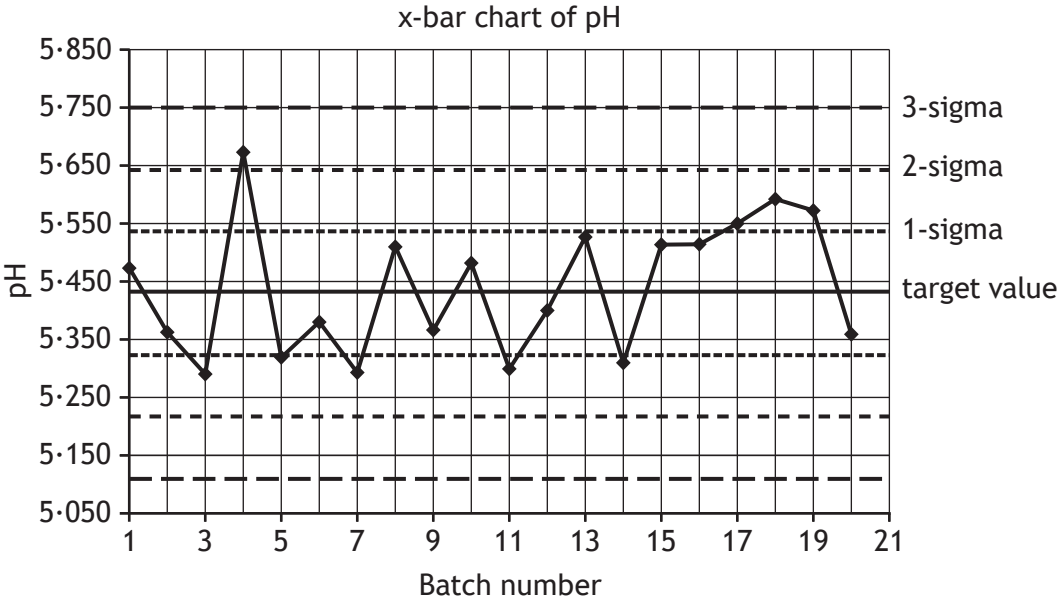
[Turn over

MARKS

9. In producing a new product, a pharmaceutical company needed to closely monitor the pH of a reaction at a critical point in the manufacturing process.

At hourly intervals a batch of 5 pH readings were taken and these were plotted on an x-bar control chart.

The first x-bar chart is shown below with the target value and the 1, 2 and 3 sigma limits.



As this is the first time this particular process has been monitored, the first 10 batch means were used to calculate the target pH value and these are listed below.

5·488	5·382	5·306	5·681	5·339
5·395	5·303	5·528	5·387	5·491

Previously studied values of pH in a very similar manufacturing process had a variance of 0·0576 and it is assumed that this new process shows the same variance.

(a) Calculate the target value and the 1-sigma limits for the above x-bar chart. **3**

The chemical engineer in charge of the process is concerned about the last four batch readings.

The first four individual readings of pH for the 21st batch are:

5·481	5·392	5·606	5·463

(b) The 5th individual reading is above the target value.

Calculate the smallest possible value for this reading that would indicate that the process might be out of statistical control. **3**

MARKS

10. An amateur statistician is a keen player of a dice game and wants to use her statistical skills to study the game. She recognises that the outcomes of the game are so varied that a full theoretical analysis would be a lengthy undertaking, so decides to look at only a small part of the game, namely the number of turns taken to win.

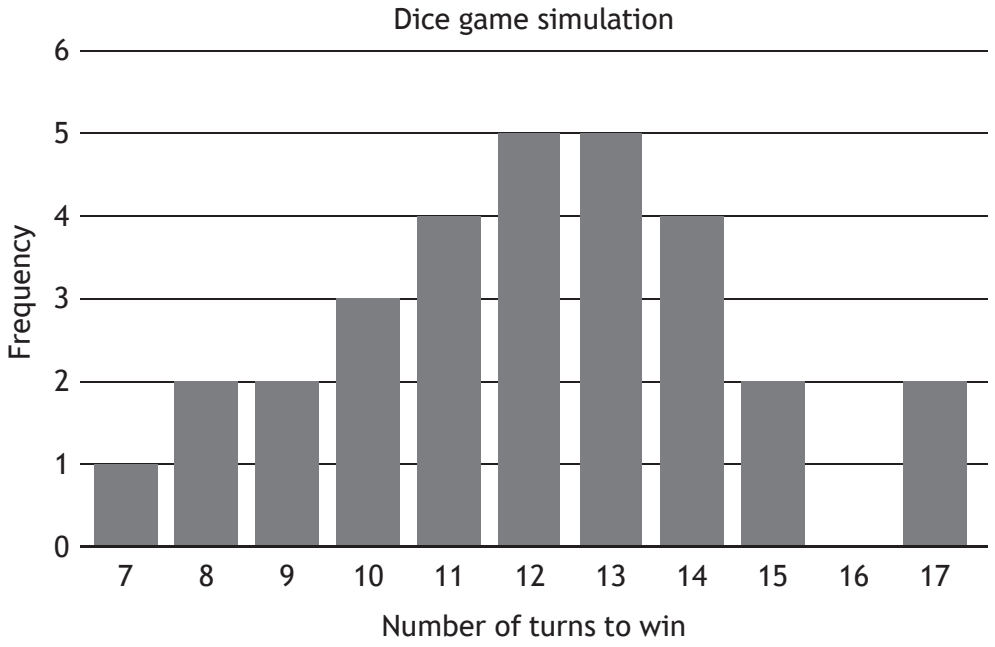
An experiment is conducted and the game is played 30 times. The number of turns taken to win, X, is recorded in the table below.

X	7	8	9	10	11	12	13	14	15	16	17
Frequency	1	2	2	3	4	5	5	4	2	0	2

(a) (i) Using the data above, construct an estimate of the probability distribution of X.

(ii) Given that $\mathrm{E}(X)=12$, show that $\mathrm{V}(X)=6$. **3**

A graph of the frequency distribution of X for the 30 games played is shown below.



The amateur statistician would like to use a theoretical model to describe X and judging from the shape of the barchart wonders if this discrete variable might be approximated by a normal distribution.

(b) (i) State, with a reason, the distribution that could be used to approximate the distribution of X.

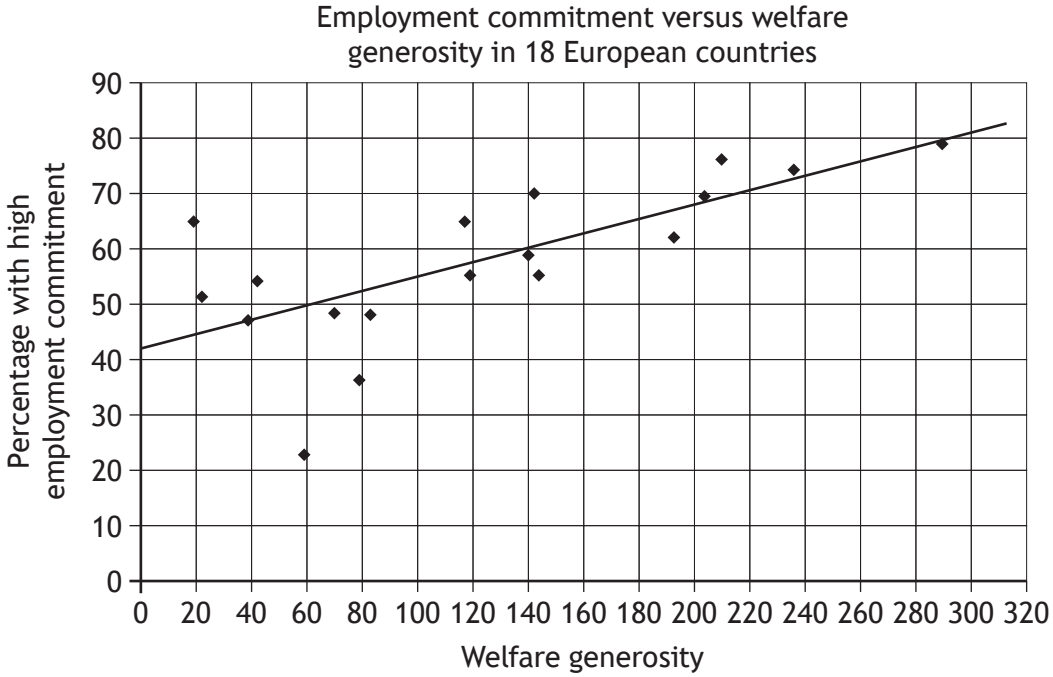
(ii) Use this approximation to calculate the probability that it takes exactly 10 turns to win.

(iii) To check the suitability of the model suggest one thing the amateur statistician might do. **6**

11. There is currently some media attention surrounding the issue of whether or not welfare benefits paid to those who are unemployed encourages people not to work. A 2015 study, conducted by two researchers at Oslo and Akershus University College of Applied Sciences in Norway, investigated how government welfare generosity, x, and people's employment commitment, y, varied across 18 European countries. An adapted version of their work is considered below.

Country	**Welfare generosity**	**Employment commitment(%)**
	x	y
Bulgaria (BG)	19	64·7
Poland (PL)	22	51·1
Estonia (EE)	39	47·0
Hungary (HU)	42	54·0
Czech Republic (CZ)	59	22·6
Portugal (PT)	70	48·3
Slovenia (SI)	79	36·2
Spain (ES)	83	47·9
Belgium (BE)	117	64·7
France (FR)	119	54·9
United Kingdom (UK)	140	58·6
Germany (DE)	142	69·9
Finland (FI)	144	54·9
Sweden (SE)	193	61·9
Switzerland (CH)	204	69·3
Netherlands (NL)	210	76·0
Denmark (DK)	236	74·1
Norway (NO)	290	78·7

11. (continued)



(a) (i) Assuming that a linear model is appropriate for this data, identify the country which might be an outlier.

(ii) The scatterplot indicates that there may be a degree of correlation between welfare spending and employment commitment.

Explain why it would be incorrect to conclude that increased welfare spending encourages people not to work.

(iii) State the feature of the data that indicates that fitting a linear model by least squares might not be appropriate. **3**

During the data checking process it is found that the employment commitment value for the Czech Republic is 42·6%, as opposed to the previously reported 22·6%.

The summary calculated statistics with this revised value of 42·6 are given below.

$$\sum x = 2208,\ \sum y = 1054{\cdot}8,\ \sum x^2 = 376752,\ \sum y^2 = 64303{\cdot}48,\ \sum xy = 141677{\cdot}3$$

$$S_{xx} = 105904,\ S_{yy} = 2492{\cdot}2$$

(b) Calculate the equation of the least squares regression line from these revised statistics and hence calculate the residual for the Czech Republic. **6**

[Turn over for next question

MARKS

12. A batch of young fish from a salmon nursery on the west coast is transported to a fish farm in two identical tanks. The fish from one tank are transferred to an empty cage at a preferred location A and the fish from the other tank to a similar empty cage at another location B.

After a year, a random sample of fish is taken from each location on the same day.

The weights (lb) of the fish are given below, together with appropriate statistics.

Location A: 13·9 11·0 11·3 12·5 13·4 11·5 14·5 13·4 12·2 11·3

$$n_a = 10,\ \bar{x}_a = 12{\cdot}5,\ s_a^2 = 1{\cdot}53$$

Location B: 14·4 12·2 17·4 12·0 16·1 10·9 19·4 10·3

A researcher decides to conduct a t-test on the data.

(a) (i) Perform this test to assess whether or not there is any evidence of a difference in mean weights from the two locations, stating the two assumptions that have been made. **10**

(ii) By considering the sample statistics, comment on why one of these assumptions may be invalid. **1**

Another researcher points out that weights of farmed salmon populations are known to have variance of 2·25 lb.

(b) Using this information, perform another test to re-assess the evidence. **3**

A third researcher questions the assumptions for conducting the t-test and is also concerned that the population variance of 2·25 cannot be known with any certainty.

(c) To overcome these concerns, state another statistical test which might be appropriate, together with its underlying assumption. **2**

[END OF QUESTION PAPER]

[BLANK PAGE]

DO NOT WRITE ON THIS PAGE

ACKNOWLEDGEMENTS

Question 8 – Data is taken from STATS19, by the Department of Transport. Contains public sector information licensed under the Open Government Licence v.3.0.

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.

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

ENTER NUMBER OF QUESTION

DO NOT WRITE IN THIS MARGIN

For Marker's Use		
Question No	Marks/Grades	