National Qualifications 2025

X803/77/12

Statistics Paper 2

WEDNESDAY, 28 MAY

10:20 AM – 1:05 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.

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 — 90

Attempt ALL questions

1. The random variable X is the score shown on a biased cubical (6-sided) die with the following probability distribution.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{60}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{19}{60}$

(a) Calculate the mean and variance of X. **3**

The random variable Y is the score shown on an unbiased octahedral (8-sided) die and has probability distribution U(8).

(b) Stating a required assumption, calculate $SD(X - Y)$. **4**

2. A shopping mall contains a large number of retail and food outlets, with an extensive car park. A local transport initiative requires information about the distance that customers are travelling to visit the shopping mall.

A random sample of 15 customers visiting the mall's food court are asked for their home postcode, and the distance from the shopping mall was calculated. The results (to the nearest km) are recorded in the list below, along with some summary statistics.

1 4 4 7 7 8 9 13 15 20 21 22 25 27 46

$Q_1 = 7$ $Q_3 = 22$ IQR = 15

(a) By calculating the upper and lower fences, comment on the highest and lowest values from this sample. **3**

It is believed that customers may travel further for specific stores. In another study, a random sample of 100 customers was taken. The customers were asked how far they had travelled, and whether they were visiting a specific retail or food outlet. The information is displayed in the table below.

	Less than 20 km	**More than 20 km**
Retail	36	26
Food	33	5

(b) Perform a chi-squared test to determine if there is an association between the type of outlet visited and the distance travelled. **6**

MARKS

3. Problems with equipment during the production process at a factory can cause a number of work stoppages which can be modelled by a Poisson distribution with a mean of 14 per month.

Using an appropriate approximation, estimate the probability that in a randomly chosen month there were more than 20 work stoppages due to equipment problems. **4**

4. The manager of a restaurant chain in America calculates the tip-per-customer for each of their restaurants by dividing the total sum of all the tips by the total number of customers. They find that from all their restaurants, the tip-per-customer has a mean of \$4.70 and a standard deviation of \$2.80.

(a) By considering the values of the mean and standard deviation, explain why the distribution of tip-per-customer is unlikely to follow a normal distribution. **1**

Over the next year, the restaurant chain expanded rapidly by opening many more new restaurants. From the new restaurants that opened in that year, a random sample of 50 restaurants was chosen and the tip-per-customer was calculated for each restaurant.

Assume that the mean and standard deviation of the tip-per-customer in the sampled restaurants is the same as those for the initial restaurants.

(b) (i) Use the central limit theorem to calculate the approximate probability that these 50 restaurants will have a mean tip-per-customer of greater than \$5.50, stating the distribution used. **3**

(ii) Describe two features of this study's data set that support the central limit theorem being used in (b) (i). **2**

[Turn over

MARKS

5. The data below are a random sample of size nine from a distribution whose median is believed to be 40.

40.1 40.3 39.5 40.6 40.8 39.9 40.0 41.3 38.3

Perform a Wilcoxon signed rank test to determine whether this sample provides evidence at the 5% level of significance that the population median is different from 40, using the hypotheses given below.

H_0: population median $= 40$

H_1: population median $\neq 40$ **6**

6. Climate change on planet Earth is predicted to lead to melting ice caps and a subsequent increase in sea level. In turn, this may have an impact upon the extreme tidal range, which is the difference between the depth of water at the highest tide and the lowest tide, over a defined period of time.

The British Oceanographic Data Centre has recording stations around the UK that include the Scottish ports of Stornoway, Tobermory, Ullapool, Port Ellen, Portpatrick, Lerwick, Aberdeen, Leith and Wick.

The monthly extreme tidal range (in metres) were recorded at these nine seaports around the coastline of Scotland for nearly every month from 2000 to 2018, and the summary data is given below.

years	mean	standard deviation	n
2000 to 2009	3.9868	1.3396	990
2010 to 2018	4.2510	1.2342	735

The following assumptions can be made:

- The recordings at each location are independent of one another.
- The nine seaports are a random sample of Scottish coastal locations.
- The monthly extreme tidal ranges are normally distributed.
- The sample standard deviations are representative of the population standard deviations.

Perform a two-sample z-test to determine at the 0.5% level whether the mean monthly extreme tidal range around the coastline of Scotland has increased over time. **5**

MARKS

7. Write down a distribution with its parameters for modelling each of the following random variables, and write down appropriate justifications for each distribution.

(a) $X =$ the number of fours obtained when rolling a fair cubical die 10 times. **3**

(The die has numbers 1, 2, 3, 4, 5 and 6 on it.)

(b) $L =$ the length of the **leftover** string after all possible 8.0 cm pieces have been precisely cut from a ball of string of unknown length. **3**

8. (a) Describe theoretically how you would obtain a simple random sample of n individuals from a larger population of size N. **3**

There are two driving test centres, A and B, in a town. Over the past year there has been mounting criticism of a lower pass rate of learner drivers at centre A.

A random sample of 29 candidates is taken from centre A, 14 of whom passed their driving test. The following month, a random sample of 30 candidates is taken from centre B, 21 of whom passed their driving test.

(b) Perform an appropriate hypothesis test at the 5% level to determine if there is evidence to support the mounting criticism of a lower pass rate at centre A, checking the conditions required for the validity of the test. **9**

(c) State one aspect of the data collection that would put the conclusion from part (b) in doubt. **1**

9. A timber company produces planks of wood of a required length. The error in the length of one plank of wood is defined as $X \sim \text{N}\left(0.1,\, 0.25^2\right)$. The error is measured in inches, where 1 inch = 2.54 centimetres.

One of the company's clients is using centimetres for a particular project and would like all measurements converted from inches to centimetres.

Let Y be the error in the length of a plank of wood when measured in centimetres.

(a) Using the approximate conversion of $Y = 2.54X$, find the mean and variance of Y. **2**

(b) Out of a batch of 80 planks, calculate the expected number that are within 1 centimetre of the required length. **4**

[Turn over

MARKS

10. A consumer regularly purchases a brand of powdered milk, which is sold in large packets with a measuring spoon. The manufacturer changes the recipe of the powdered milk.

For the original recipe, the manufacturer's instructions indicated that one level spoonful using the provided spoon had a mean mass of 4.5 grams. Following the change of recipe, the consumer questions whether the mean mass of powdered milk per spoonful has changed.

The consumer measures six spoonfuls of new recipe powdered milk and weighs them using high-precision weighing scales. Taking the random variable X to be the mass of a spoonful of powdered milk (in grams), the consumer generates the following results:

$$\sum x = 27.9 \text{ and } \sum x^2 = 130.2$$

It is assumed that X is normally distributed.

(a) Stating one further assumption, perform a parametric test at the 10% level to assess whether the new recipe gives a different mean mass per spoonful of powdered milk to that of the original recipe. **7**

The manufacturer carries out a similar investigation, using a population variance taken from previous large-scale research.

(b) Comment on how the choice of hypothesis test carried out by the manufacturer would differ to that of the consumer in part (a). **1**

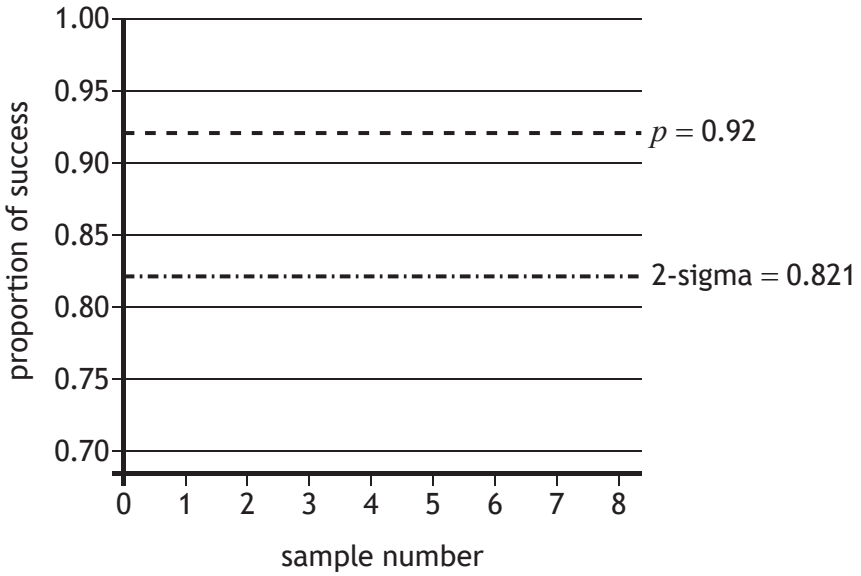
MARKS

11. The Customer Service department of a large retail chain has a history of resolving 92% of phone enquiries within seven minutes. During the summer months, experienced phone operators go on holiday and need to be replaced by temporary staff.

To maintain the department's high standards during the summer months, the Customer Service Manager monitored the time taken for staff to resolve an enquiry by taking a random sample of calls every day over several days. A phone enquiry was considered to be a success if a phone operator resolved a customer enquiry within 7 minutes, and a failure if a phone operator took longer than 7 minutes.

The outline of a p-chart is shown below, with values shown for the historical proportion of success, p, and the lower 2-sigma limit.

***p*-chart for success in resolving enquiry within 7 minutes**



(a) (i) Calculate the lower 1-sigma and lower 3-sigma limits. **3**

(ii) Without calculating any of the upper limits, give a reason why the Customer Service Manager does not need to consider them when monitoring the resolution times. **1**

Sample 1 and sample 2 gave success proportions of 0.800 and 0.833 respectively.

(b) If sample 3 gave a success proportion of 0.815, determine with justification if the Customer Service Manager should be concerned. **2**

[Turn over

MARKS

12. The bounce height of a tennis ball is tested by dropping it vertically from a set height onto a concrete surface and measuring the height of the ball's first bounce.

A tennis ball manufacturing company tested a sample of 15 balls and found that the mean bounce height was 139.5 cm with standard deviation 0.7 cm.

(a) For the given sample, calculate an approximate 98% confidence interval for the mean bounce height, stating the two assumptions required. **5**

A tennis tournament requires the mean bounce height of the balls to be at least 141 cm.

(b) Comment on whether or not you would recommend the batch of balls from which the sample was drawn, to be used in the tournament. **2**

13. Two events A and B are such that:

$P(A) = 0.69 \qquad P(A \cap \overline{B}) = 0.55 \qquad P(\overline{A} \cap \overline{B}) = 0.06$

By using either a Venn diagram or a tree diagram, or otherwise, calculate:

(a) $P(\overline{A} \cap B)$ **2**

(b) $P((A \cap B) \cup (\overline{A} \cap \overline{B}))$ **2**

(c) $P(A \mid B)$ **3**

[END OF QUESTION PAPER]