AH National Qualifications 2023 MODIFIED

X803/77/12

Statistics Paper 2

FRIDAY, 19 MAY

10:30 AM – 1:00 PM

Total marks — 80

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.

X8037712

A/HTP

MARKS

Total marks — 80

Attempt ALL questions

1. The heights of boys aged five years are normally distributed with mean 109 cm and standard deviation 7 cm.

(a) A boy aged five years is chosen at random. Calculate the probability that he is taller than 111 cm. **2**

(b) A random sample of 25 boys aged five years is chosen. State the sampling distribution of the sample mean and hence calculate the probability that the sample mean is greater than 111 cm. **3**

(c) By considering the sampling distribution of the sample mean and its spread, explain why the answers to parts (a) and (b) are not the same. **1**

2. Duncan suspected the step-counter on his mobile phone was over-counting the number of steps he took when he was walking in his local area. On one such walk he counted a series of ten random sets of 300 steps and then recorded after each set the number shown by the step-counter on his phone. The results he recorded are shown below.

320 310 321 304 298 328 296 307 314 295

By stating a required assumption, conduct a Wilcoxon Signed Rank test at the 5% level of significance, to determine whether there is evidence that the step counter from the mobile phone over-counts the median number of steps that Duncan takes. **6**

3. The table below shows the historical proportions of the Scottish population belonging to each of the different blood groups. There are four main blood groups O, A, B and AB and each can be either positive (+) or negative (–).

Blood Group	O+	O−	A+	A−	B+	B−	AB+	AB−
Proportion	0.409	0.095	0.288	0.063	0.092	0.020	0.027	0.006

People who consent to give blood are called blood donors.

(a) Calculate the probability that in a random sample of 20 Scottish blood donors, at least two will be blood group B−. **3**

(b) Using a suitable approximation with justification, estimate the probability that, in a random sample of 50 Scottish blood donors, at most 30 will be blood group O+ or O−. **6**

MARKS

4. Genetic theory dictates that a certain plant should have its three types of offspring in the ratio $1:1:2$. A random sample of 320 such offspring yielded frequencies of 78, 90 and 152 respectively.

A chi-squared goodness-of-fit test is performed using the following hypotheses:

H_0: data follows the specified ratio

H_1: data does not follow the specified ratio

Show, by using a chi-squared goodness-of-fit test, that the random sample provides evidence at the 10% level to support the genetic theory. **4**

5. The discrete random variable X takes the values 0, 1, 2, 3 or 4 with the following probability distribution.

x	0	1	2	3	4
$\mathrm{P}(X = x)$	p	p	$2p$	$5p$	

(a) (i) Find $\mathrm{P}(X = 4)$ and hence show that $\mathrm{E}(X) = 4 - 16p$. **2**

(ii) If $\mathrm{E}(X) = 3$, determine the value of p and hence calculate $\mathrm{V}(X)$. **4**

A new random variable K is defined as $K = 2Y - X + 3$, where $Y \sim \mathrm{Po}(1)$, X is defined as above, and X and Y are independent.

(b) Calculate the mean and standard deviation of K. **4**

[Turn over

MARKS

6. At birth, a newborn baby's length is measured from the top of their head to the bottom of one of their heels. This measurement is taken instead of height, as it is easier to establish with a suitable degree of accuracy. The length at birth for a full-term baby is believed to be normally distributed with mean 50 cm.

A midwife has a theory that the mean lengths of babies born to male basketball players are greater than that of the general population, due to the height of the father.

A random sample of 75 full-term babies born to basketball players is taken, where the father had a height of at least 2 metres. The random variable X is the length of the baby at birth, in centimetres. The following statistics are obtained.

$$\sum x = 3840 \quad \sum x^2 = 198240$$

Assume the sample standard deviation is a good estimate of the population standard deviation. Perform a suitable test, at the 1% level of significance, and comment on the midwife's theory. **6**

7. A packed lunch consists of a sandwich, a drink and a piece of fruit. Each part of the packed lunch is randomly selected from the available options and the selection of sandwich, drink and piece of fruit are all independent of each other.

The filling in the sandwich has a 50% chance of being jam, a 30% chance of being cheese and a 20% chance of being tuna. The drink is either water or lemonade and the piece of fruit is either an apple or a banana.

The probability of a packed lunch containing a tuna sandwich and water is 0.035.

(a) Calculate the probability of having water to drink, given that the sandwich filling is tuna. **2**

The probability of a packed lunch containing a cheese sandwich and a banana is 0.12.

(b) Calculate the probability of the packed lunch having both a jam sandwich and an apple. **5**

MARKS

8. Plasma is the pale yellow liquid component of blood and makes up more than half of the body's total blood volume in healthy people.

For a random sample of 25 healthy women, body mass (kilograms) and plasma volume (litres) were determined and a scatterplot of the data indicated that a linear association appeared to be present.

The product moment correlation coefficient was found to be 0.652.

Perform a hypothesis test at the 0.1% level on the linear association between body mass and plasma volume in healthy women. **6**

[Turn over

9. A fitness tracker is a device for monitoring and tracking various fitness-related measurements such as distance travelled, heart rate, or calorie consumption. A sports scientist investigated whether wearing a fitness tracker encouraged runners to run further.

A randomly selected group of 12 runners were given fitness trackers that alerted them when they reached a target distance, specified by the runner. They were asked to run three times per week for two weeks — one week with the tracker and one without. The allocation of which week each runner had a tracker was randomised. During the runs with the tracker, the runner was alerted when they reached their target distance. On the runs without the tracker, they received no information about their progress.

The table below shows the mean distance covered by each runner, in kilometres.

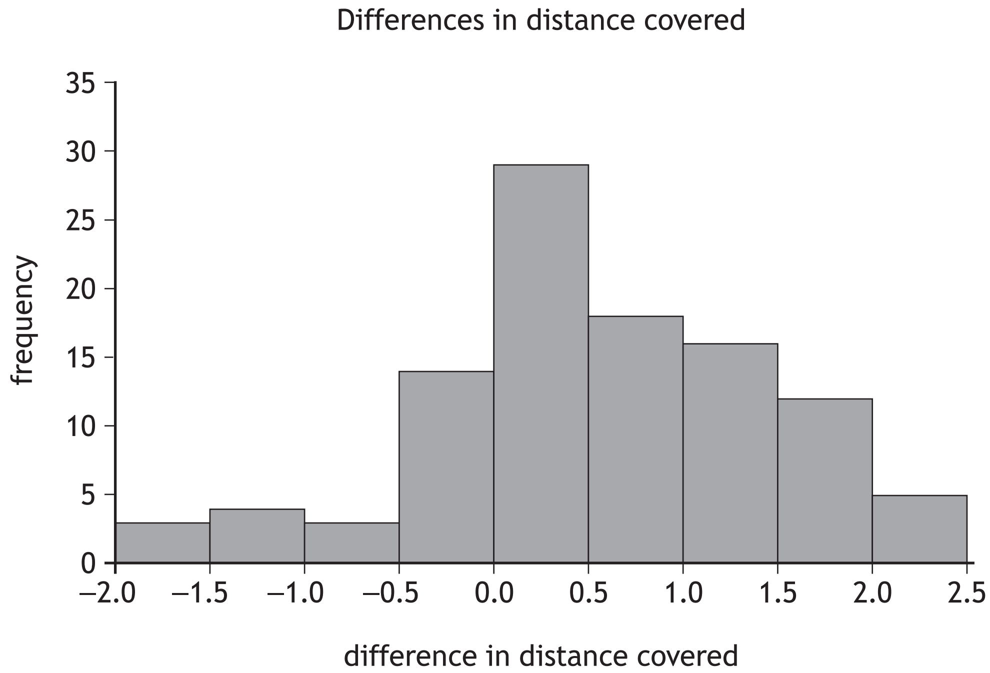
Runner	With tracker	Without tracker	Difference
1	5.1	4	1.1
2	10	9.5	0.5
3	10.8	12	–1.2
4	7.5	5.5	2
5	6.2	5.9	0.3
6	10.2	11	–0.8
7	5.4	4.8	0.6
8	4.2	3.5	0.7
9	8.1	6.5	1.6
10	11.1	11.5	–0.4
11	10.2	9.4	0.8
12	5.3	5.1	0.2

The **Difference** (= With tracker – Without tracker) in the distance covered has mean 0.45 and standard deviation 0.927.

(a) Assuming that the differences are normally distributed, perform a suitable parametric test to determine if runners ran further when wearing a fitness tracker. **6**

9. (continued)

To collect further data, the sports scientist extended the investigation to a random sample of 104 runners, and used a histogram to chart the difference between the runs with and without the fitness tracker.



(b) (i) Comment on the assumption of normality of the differences, with reference to the histogram. **1**

(ii) Name a non-parametric test that could also be considered for use on this data and comment on its suitability. **2**

[Turn over

MARKS

10. The American Department for Housing and Urban Development (HUD) publishes annual reports on the number of homeless people across all 55 American states and major territories. One section of their report records the estimated number of ex-army veterans who are homeless and how many of these homeless veterans are in sheltered accommodation.

Over the 8 year period from 2010 to 2017, the mean proportion of homeless veterans in sheltered accommodation was 62.4%.

In 2018, the report stated that of 37 878 homeless veterans, 23 312 were in sheltered accommodation.

Perform an appropriate hypothesis test to determine if the HUD report provides any evidence, at the 0.5% significance level, that the proportion of homeless veterans in sheltered accommodation in 2018 is significantly different from the previous 8 years' data. **6**

11. A random variable, X, is normally distributed with mean μ and variance σ^2.

It is known that the probability that X is greater than 24 is 0.05.

The probability that X is less than 17 is 0.1.

Calculate the values of μ and σ. **4**

12. At a polling station in a recent election in Scotland, data from a random sample of 100 voters showed that 55% of them were in favour of a certain candidate being elected.

(a) Construct an approximate 99% confidence interval for the proportion of the population who were in favour of this candidate being elected, and explain why it is only an approximate interval. **4**

For a candidate to be elected, they require to win 50% of the votes cast.

(b) Assuming the same sample proportion is in favour, calculate the smallest size of sample that should have been taken in order for a 99% confidence interval to indicate that the candidate would be elected. **3**

[END OF QUESTION PAPER]