



National
Qualifications
2022 MODIFIED

X803/77/12

**Statistics
Paper 2**

MONDAY, 23 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.



* X 8 0 3 7 7 1 2 *

Total marks — 80
Attempt ALL questions

1. Data was collected by a group of university students investigating the prevalence of a parasite infection on fish and the observations are recorded in a contingency table shown below.

		Sex of fish		Total
		Male	Female	
Infected	Yes	76	129	205
	No	399	332	731
Total		475	461	936

The students wish to determine if there is an association between the prevalence of infection and the sex of the fish.

Carry out a chi-squared hypothesis test, at the 1% level of significance, to determine if there is any evidence of such an association.

6

2. An apiary consists of two beehives, A and B. The number of bees leaving the hives every one-minute-interval is recorded. It is assumed that bees leave the hives independently of each other. The bees from hive A leave at a mean rate of 2.3 per minute and those from hive B at a mean rate of 1.7 per minute.

The random variables X and Y are defined as the number of bees leaving hive A and hive B respectively.

- (a) State the distribution of X .

1

During any given minute, calculate the probability that:

- (b) no bees leave hive A
 (c) exactly 2 bees leave each hive
 (d) more than 5 bees leave the apiary.

1

2

3

3. In a game, 5 cards show the numbers 0, 0, 2, 4 and 4.

Players select two cards at random, without replacement and their score T is the total of the numbers on their two cards.

Copy and complete the following probability distribution table for T and calculate $E(T)$ and $V(T)$.

4

t	0	2	4	6	8
$P(T = t)$					

4. Figures from the Driver and Vehicle Licensing Agency show that in a large Scottish city 44% of driving tests are passed.

A random sample of 104 test results is selected.

- (a) Calculate the probability that exactly half of these tests were passed.

3

- (b) Using an approximation, determine the probability that between 40 and 50 (inclusive) of these tests were passed.

4

5. The prelim marks of a random sample of nine students who studied both French and German at Advanced Higher are shown in the table.

	Adam	Mike	Rob	Emma	Jack	Keri	Elena	Seif	Jess
French	67	83	71	59	49	89	42	55	77
German	64	82	71	62	42	85	39	50	75

- (a) State why it would not be appropriate to conduct a t -test for a difference in population means on this data.

1

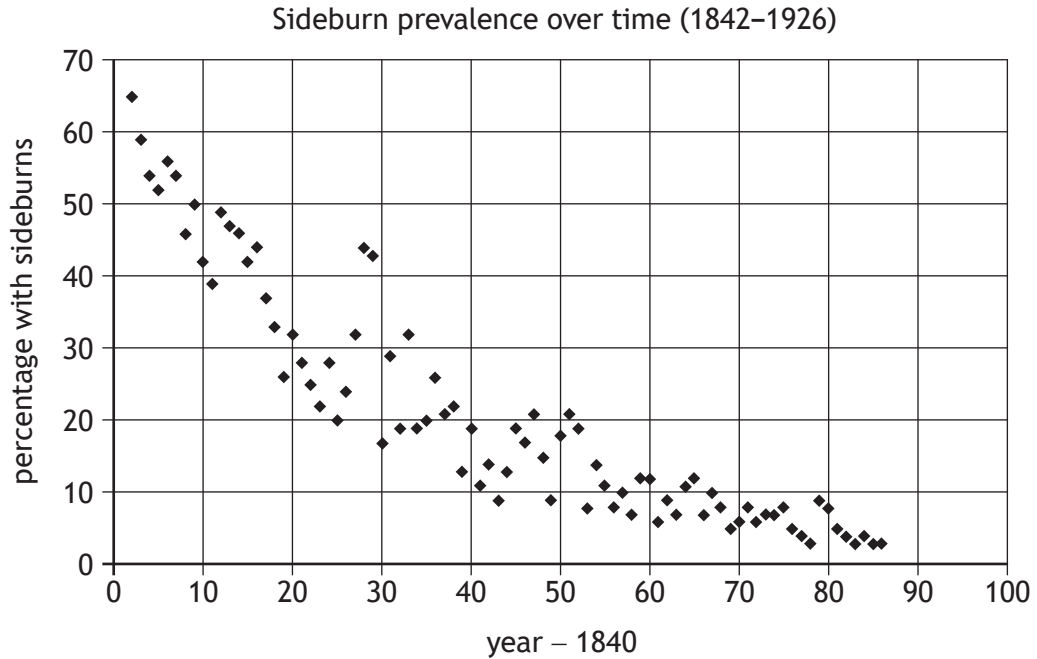
- (b) Stating an appropriate assumption, perform a parametric test to investigate if there is a difference between French and German prelim marks.

8

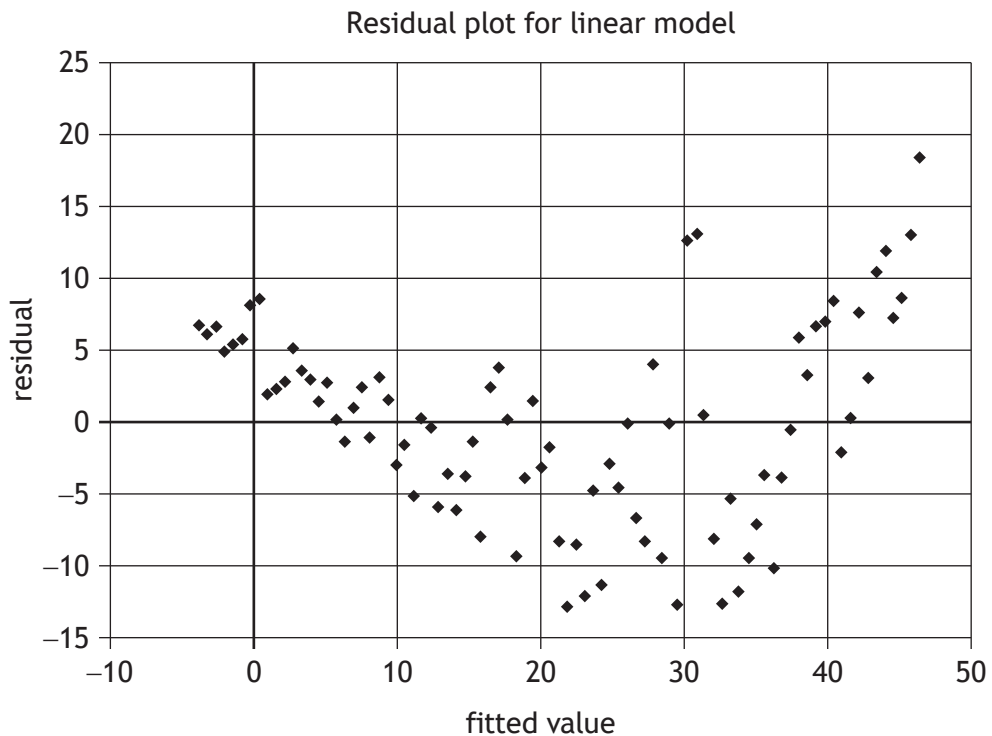
[Turn over

6. Writing in the *American Journal of Sociology* in 1976, Dwight E. Richardson investigated 'Fashions in Shaving and Trimming of the Beard between 1842 and 1926'. Consideration is given below to an adapted sub-section of his work that looked solely at the popularity of sideburns.

For this scatterplot, year is plotted on the x -axis (with $x = \text{year} - 1840$). Richardson studied pictures in the weekly pictorial magazine *Illustrated London News* to determine the percentage of men, y , in each magazine that had sideburns and no other facial hair.



A least squares linear regression line is used to model these data and the corresponding residual plot is shown below.



6. (continued)

- (a) Describe two characteristics of the pattern of points on the residual plot that indicates the linear model is not appropriate.

2

To improve the linear model, $\log_{10} y$ is plotted against x and this gives a second model, $\log_{10} y = a + bx$. The following statistics are calculated, with $w = \log_{10} y$.

$$\sum x = 3740, \quad \sum w = 101.2529, \quad n = 85$$

$$S_{xx} = 51170, \quad S_{ww} = 11.145, \quad S_{xw} = -715.456$$

- (b) Determine the least squares regression equation of w on x , and use it to predict the percentage of men with sideburns and no other facial hair in 1927.

4

(Note: $\log_{10} p = q$ can be written as $10^q = p$)

7. The estimator for the population proportion, p , from a sample proportion is $\frac{X}{n}$, where X is the number of 'successes', and $X \sim B(n, p)$.

- (a) Use the laws of expectation and variance to show that

$$E\left(\frac{X}{n}\right) = p \text{ and } V\left(\frac{X}{n}\right) = \frac{pq}{n}, \text{ where } q = 1 - p.$$

2

In an experiment, there were 14 successes from 50 independent trials.

- (b) Calculate an approximate 99% confidence interval for the population proportion of successes, checking the validity of the method used.

4

[Turn over

8. A children's game consists of an unbiased 5-sided spinner and a pack of 32 cards.
- The spinner has numbers 1, 2, 3, 4 and 5.
 - The cards have one of the numbers 1, 2, 3, 4 or 5 displayed on one side, and a goldfish or shark on the other.

To play the game, all the cards are first shuffled and laid out in front of the players with the card numbers facing upwards.

Players use the unbiased 5-sided spinner to determine which numbered card they will select **before** turning it over to reveal either a goldfish or a shark.

The player loses if a shark is revealed.

The distribution of the cards is as follows:

Card number	Goldfish cards	Shark cards
1	3	2
2	7	0
3	6	0
4	5	3
5	6	0

- (a) Calculate the probability of spinning a number 4 and then revealing a goldfish. 2
- (b) Calculate the probability of:
- (i) finding a shark given that you spun a 1 1
 - (ii) losing the game 3
 - (iii) spinning a number 1, given that you have lost the game. 3

9. (a) Describe what the central limit theorem states, for sufficiently large samples from a population.

2

Wooden battens are sold as having a mean width of 50 mm. A friend of yours is finding difficulty in fitting such battens into larger wooden structures and asks you to investigate the problem. You obtain a random sample of 45 such battens and you calculate their mean width to be 52.6 mm with variance 103.25 mm².

- (b) Using the central limit theorem and a 5% significance level, perform a z -test on the following hypotheses, stating a further assumption that is required.

5

$$H_0: \text{mean batten width} = 50 \text{ mm}$$

$$H_1: \text{mean batten width} > 50 \text{ mm}$$

10. After a radioactive leak from a nuclear power station, each of 6 surrounding towns was given an exposure index, x . The impact of the leak was monitored and ten years later, the number of deaths per 10 000 of the population, y , for each of the 6 towns was recorded. The following statistics were calculated:

$$\sum x = 64.40 \quad S_{xx} = 278.61 \quad S_{yy} = 10.95 \quad S_{xy} = 46.29 \quad n = 6$$

- (a) Conduct an appropriate test at the 5% significance level to investigate whether the exposure index and the number of deaths per 10 000 are linearly associated, stating any assumption required.
- (b) A journalist concludes from the analysis that the exposure to the leak caused the deaths. Suggest why this may not be a valid conclusion.

7

1

[Turn over

11. A group of zoology students studying animal behaviour designed an experiment to investigate the impact of noise on fennec foxes. The reaction times (seconds) to loud music were recorded for two independent random samples of 10 adult and 10 juvenile foxes, and are shown below.

Adults	1.3	2.2	1.5	3.5	0.3	2.7	3.5	2.3	2.9	4.0
Juveniles	1.1	4.1	1.7	1.3	0.7	1.9	2.9	1.1	2.8	0.9

- (a) (i) Construct a back-to-back stem-and-leaf diagram to compare the reaction times of adult and juvenile foxes, and comment on what the diagram shows. 2
- (ii) Perform a Mann-Whitney test, using the juvenile rank sum of 89, to determine if there is a difference in median reaction times between adult and juvenile foxes. 4

From larger and longer term studies, models of the reaction times of adult fennec foxes follow $N(2.5, 0.5)$ and those of the juveniles follow $N(2.0, 0.3)$.

- (b) Calculate the probability that the reaction time of a randomly chosen adult is slower than that of a juvenile fox. 5

[END OF QUESTION PAPER]