

Paper Reference 9ST0/02
Pearson Edexcel Level 3 GCE

Statistics

Advanced

PAPER 2: Statistical Inference

Wednesday 12 June 2024 – Afternoon

Time: 2 hours

**DO NOT RETURN THIS QUESTION PAPER
WITH THE ANSWER BOOKLET AT THE END OF
THE EXAMINATION.**

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have retrievable mathematical formulae stored in them.

YOU MUST HAVE

Statistical formulae and tables booklet, calculator

YOU WILL BE GIVEN

Data Booklet

Answer Booklet

INSTRUCTIONS

In the boxes on the Answer Booklet and on the Data Booklet, write your name, centre number and candidate number.

Answer ALL questions and ensure that your answers to parts of questions are clearly labelled.

Answer the questions in the Answer Booklet or on the separate data sheets – there may be more space than you need.

Do NOT write on this Question Paper.

You should show sufficient working to make your methods clear. Answers without working may not gain full credit.

Unless otherwise stated, inexact answers should be given to three significant figures.

Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

INFORMATION

A booklet ‘Statistical formulae and tables’ is provided.

**There are 7 questions in this Question Paper.
The total mark for this paper is 80.**

**The marks for EACH question are shown in brackets
– use this as a guide as to how much time to spend on
each question.**

**There may be spare copies of some data sheets in case
you need them.**

ADVICE

**Read each question carefully before you start to answer
it.**

Try to answer every question.

Check your answers if you have time at the end.

**If you change your mind about an answer, cross it out
and put your new answer and any working underneath.**

Answer ALL questions. Write your answers in the Answer Booklet or on the separate data sheets.

- 1. Refer to the diagram and the table for Question 1 in the Data Booklet.**

Brenda is a yoga teacher.

She is researching the benefits of regular yoga practice on reducing eye fatigue.

To carry out her research, Brenda seeks 12 volunteer students, who have never practised yoga before, to join her yoga classes.

She asks each student to complete a questionnaire BEFORE they start her classes.

In her questionnaire, the first question is shown in the diagram in the Data Booklet.

(continued on the next page)

1. continued.

After attending her classes for three months, these 12 students are asked again to complete the same questionnaire.

Both scores for the first question, for each of these 12 students, are given in the table in the Data Booklet.

Brenda believes that regular yoga practice can reduce the average eye fatigue score.

(a) Making any necessary assumptions, carry out a paired t -test to investigate whether there is evidence to support Brenda's belief.

(7 marks)

(b) Give TWO possible sources of bias in this experiment.

(2 marks)

(Total for Question 1 is 9 marks)

2. Refer to the information for Question 2 in the Data Booklet.

It shows the details of a survey carried out by Briony.

Using the results of Briony's survey carry out a hypothesis test, using a **distributional approximation**, to investigate whether the percentage of those aged under 18 who agree with the statement is greater than the percentage of adults in Great Britain who agree with the statement.

You may assume that responses were independent of one another.

(Total for Question 2 is 6 marks)

- 3. Refer to Table 1, Table 2 and Table 3 for Question 3 in the Data Booklet.**

Adya, a retail researcher, wants to investigate whether there is a relationship between ages of members of the British public and their preferred supermarket.

Adya decides to post on social media asking members of the public living in her town to complete a survey.

Her survey asks each person to categorise themselves as ‘student’, ‘young adult’, ‘older adult’ or ‘retired’ and to select their preferred supermarket A, B, C, D or E

Her summarised data is given in Table 1.

(continued on the next page)

3. continued.

Table 2 in the Data Booklet shows an incomplete table of the expected frequencies.

Adya decides to perform a test for association on this data and calculates some of the expected frequencies, using the observed frequencies in Table 1 in the Data Booklet.

- (a) Complete the remaining expected frequencies, correct to one decimal place in Table 2.**

There are eight spaces to fill.

(2 marks)

After calculating the expected frequencies, Adya combined the data, in Table 1 and Table 2, for supermarkets **C, **D** and **E** into a new category “Other” in order to make the test of association valid.**

- (b) Explain why Adya needed to do this.**

(1 mark)

(continued on the next page)

Turn over

3. continued.

Her combined observed frequencies are given in Table 3 in the Data Booklet.

(c) Making any necessary assumptions, carry out a test using the data in Table 2 and Table 3 to investigate whether there is any association between age and preferred supermarket.

(8 marks)

(d) Suggest TWO improvements that Adya could make to this investigation in order to reduce bias.

For each improvement, explain how it could lead to reduced bias.

(4 marks)

(Total for Question 3 is 15 marks)

4. Refer to the information for Question 4 in the Data Booklet.

It shows the details of a survey carried out by the Office for National Statistics.

- (a) Conduct a hypothesis test to investigate whether the proportion of **larger businesses** in Northern Ireland using or intending to use increased homeworking was more than the proportion of **smaller businesses** in Northern Ireland using or intending to use this approach.

(8 marks)

- (b) Briefly summarise your conclusion to the hypothesis test in a manner suitable for publishing in a newspaper article.

(2 marks)

(continued on the next page)

4. continued.

(c) State ONE assumption you have made with regards to the sample when carrying out the hypothesis test.

(1 mark)

(d) Comment on the likely validity of your conclusion in (a) for businesses across the UK.

(1 mark)

(Total for Question 4 is 12 marks)

5. Miku is interested in testing whether participants skydiving for the first time will experience a higher level of stress than those who had previously been skydiving.

Cortisol is a hormone released by humans when experiencing stressful events and situations and can be measured by either taking a sweat or saliva sample.

Miku chooses to measure the amount of cortisol in sweat samples of 13 first-time skydivers prior to their first jump.

The amounts of cortisol measured, in $\mu\text{g/dL}$, are shown below.

4.20	4.12	3.76	3.85	3.64	4.35	3.73
4.52	3.82	4.13	4.02	4.10	3.80	

(continued on the next page)

5. continued.

From a larger study conducted on experienced skydivers it was found that their average cortisol measure prior to jumping was $3.82 \mu\text{g/dL}$

(a) Conduct a Wilcoxon signed–rank hypothesis test to investigate Miku’s claim.

State the necessary assumption that must be made about the distribution of cortisol measurements for this test to be valid.

(8 marks)

It was later discovered that the average cortisol levels for experienced skydivers had been measured using saliva samples.

(b) Based on this information, explain why the conclusion in (a) may not be valid.

(1 mark)

(continued on the next page)

5. continued.

Refer to the table for Question 5(c) in the Data Booklet.

After further research Miku, found that there are several scientific studies that had suggested a strong positive association between the values of cortisol levels in sweat and saliva samples.

Miku found, in a secondary source, cortisol measurements, obtained from both sweat and saliva, for each of 9 adults.

These are shown in the table.

Miku believes that the sample data are NOT from a bivariate normal distribution.

- (c) Using the data in the table, investigate whether there is a positive association between the cortisol levels in sweat and saliva readings.**
(6 marks)

(continued on the next page)

5. continued.

**Based on the result of the hypothesis test in (c),
Miku decides that the test in (a) is still valid.**

(d) Explain why Miku is NOT correct.

(3 marks)

(Total for Question 5 is 18 marks)

6. Refer to the information for Question 6 in the Data Booklet.

It shows the details of a six-week study.

- (a) Using the information in the table in the Data Booklet calculate the value of Cohen's d for the effect of HIIT training on VO2 max levels.

You may assume that VO2 max levels in the two groups are normally distributed and have a common variance.

(2 marks)

(continued on the next page)

6. continued.

- (b) Use the results of the t -test AND the value of Cohen's d calculated in (a) to draw conclusions about the difference in average VO₂ max levels for adults with mild asthma participating in the HIIT programme and those who are participating in their usual recommended exercise.**

(3 marks)

(Total for Question 6 is 5 marks)

7. Refer to the table for Question 7 in the Data Booklet.
In January 2023, Michael was working at a financial advisory firm in the North West of England.

He was interested in investigating whether people living in Manchester and Liverpool have different levels of savings.

He decided to use his company records for his investigation.

An extract of these records is shown in the table.

- (a) Explain how Michael could use a spreadsheet function to help SELECT appropriate data for this investigation.
(2 marks)

(continued on the next page)

7. continued.

Refer to the table for Question 7(b) in the Data Booklet.

Having selected the relevant data Michael calculated the summary statistics shown in the table for levels of savings.

(b) Making any necessary assumptions, carry out a hypothesis test to investigate whether people living in Manchester and Liverpool have different **average levels of savings.**

(8 marks)

(c) Explain

(i) why the only error that could have occurred in (b) is a Type **II error,**

(1 mark)

(ii) the meaning of this error in the context of (b)

(1 mark)

(continued on the next page)

Turn over

7. continued.

During November 2022, a large survey was carried out across the United Kingdom relating to savings.

It was found that, on average, the difference between savings per household for those in Manchester and in Liverpool was actually £270

- (d) Using the information from this large survey, calculate the probability of the possible Type II error described in (c)
(3 marks)

(Total for Question 7 is 15 marks)

TOTAL FOR PAPER IS 80 MARKS

END OF PAPER

Sources

Question 5

[Data source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4373275/>]

Question 7

[Source: Savings statistics: Average savings in the UK (finder.com)]