Please check the examination details below before entering your candidate information

| Candidate surname | Other names |
|---|---|
| | |

**Pearson Edexcel Level 3 GCE**

| Centre Number | Candidate Number |
|---|---|
| | |

**Time** 2 hours

**Paper reference** **9ST0/03**

# Statistics

## Advanced
## PAPER 3: Statistics in Practice

**You must have:**
Statistical formulae and tables booklet
Calculator

Total Marks

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

## Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with 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 spaces provided – *there may be more space than you need.*
- 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.*

## 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.
- Good luck with your examination.

*Turn over* ▶

P66657A



1/1/1/1/

P 6 6 6 5 7 A 0 1 2 8

Pearson

**Answer ALL questions. Write your answers in the spaces provided.**

**1** Surveys of whales were made in the Southern Ocean between 1995 and 2004.

**Figure 1** shows results from these surveys for the total number of sightings of individuals of three different whale species in the months of January to April.

| | | Month | | | | |
|---|---|---|---|---|---|---|
| | | **January J** | **February F** | **March M** | **April A** | **Total** |
| **Type of whale** | **Orca C** | 190 | 203 | 89 | 12 | **494** |
| | **Fin whale W** | 45 | 133 | 33 | 0 | **211** |
| | **Long-finned pilot whale L** | 234 | 150 | 240 | 100 | **724** |
| | **Total** | **469** | **486** | **362** | **112** | **1429** |

(Data source: http://ipt.obis.org/nonode/resource?r=whale_survey)

**Figure 1**

*J* denotes the event 'sighting of whale was in January'.

*F* denotes the event 'sighting of whale was in February'.

*M* denotes the event 'sighting of whale was in March'.

*A* denotes the event 'sighting of whale was in April'.

*C* denotes the event 'the whale sighted was an orca'.

*W* denotes the event 'the whale sighted was a fin whale'.

*L* denotes the event 'the whale sighted was a long-finned pilot whale'.

**Question 1 continued**

(a) For a randomly selected whale sighting find,

(i) P($L$)

**(1)**

(ii) P($A'$)

**(1)**

(iii) P($W \cap F$)

**(1)**

**Question 1 continued**

(iv) $P(C \mid M \cup A)$

**(3)**

(b) In the context of the question, explain what $P(C \mid M \cup A)$ represents.

**(1)**

**Question 1 continued**

(c) Demonstrate, with numerical justification, that the events $F$ and $W$ are **not** statistically independent.

**(2)**

**Question 1 continued**

The data in **Figure 1** is shown in **Figure 2**.

| Observed frequencies | | Month | | | | |
|---|---|---|---|---|---|---|
| | | **January** | **February** | **March** | **April** | **Total** |
| **Type of whale** | **Orca** | 190 | 203 | 89 | 12 | **494** |
| | **Fin whale** | 45 | 133 | 33 | 0 | **211** |
| | **Long-finned pilot whale** | 234 | 150 | 240 | 100 | **724** |
| | **Total** | **469** | **486** | **362** | **112** | **1429** |

**Figure 2**

Some of the calculated expected values for a $\chi^2$ test of independence are shown in **Figure 3**.

| Expected frequencies | | Month | | | | |
|---|---|---|---|---|---|---|
| | | **January** | **February** | **March** | **April** | **Total** |
| **Type of whale** | **Orca** | 162.1 | 168.0 | 125.1 | | **494** |
| | **Fin whale** | 69.3 | 71.8 | | | **211** |
| | **Long-finned pilot whale** | 237.6 | | | | **724** |
| | **Total** | **469** | **486** | **362** | **112** | **1429** |

**Figure 3**

**Question 1 continued**

(d) Complete the table in **Figure 3** and calculate the $\chi^2$ test statistic.

**(4)**

**Question 1 continued**

(e) State the degrees of freedom and the critical value for a $\chi^2$ test based on the data in **Figure 2**.

**(2)**

(f) Investigate whether the data in **Figure 2** provides sufficient evidence to conclude that the species of a whale, sighted in the Southern Ocean, is not independent of the month of the year in which it is sighted.

You must demonstrate the use of your answers to **both** parts (d) and (e).

You should assume that the samples were random and that each whale sighting is independent of the others.

Describe the nature of any association found between species of whale and month of sighting.

**(4)**

**Question 1 continued**

(g) Consider the context of the test in part (f).

State, with justification, whether the assumption that each whale sighting is independent of the others is reasonable.

**(1)**

**(Total for Question 1 is 20 marks)**

**2** Mary is a student midwife doing a project comparing births in different countries. She obtains the total fertility rates in 2000–2005 and 2010–2015 for a random sample of 11 countries.

The total fertility rate is the average number of children a woman in that country will have.

This data and the paired differences are shown in **Figure 4**.

| | **Total fertility rate** | | | | |
|---|---|---|---|---|---|
| **Country** | **2000–2005** | **2010–2015** | **Differences** | | |
| Austria | 1.38 | 1.45 | −0.07 | | |
| Indonesia | 2.53 | 2.45 | 0.08 | | |
| Panama | 2.68 | 2.60 | 0.08 | | |
| Cuba | 1.59 | 1.71 | −0.12 | | |
| Israel | 2.91 | 3.04 | −0.13 | | |
| Costa Rica | 2.15 | 1.85 | 0.30 | | |
| United Arab Emirates | 2.40 | 1.70 | 0.70 | | |
| Côte d'Ivoire | 5.68 | 4.95 | 0.73 | | |
| Guinea | 5.91 | 5.14 | 0.77 | | |
| Guatemala | 4.33 | 3.19 | 1.14 | | |
| Afghanistan | 7.18 | 5.45 | 1.73 | | |

**Figure 4**

(Data source: http://www.data.un.org)

(a) Making any necessary assumptions, perform a Wilcoxon signed-rank test to investigate whether total fertility rates by **country** changed on average between 2000–2005 and 2010–2015.

You may use the spare columns in **Figure 4** for your working if desired.

**(6)**

**Question 2 continued**

**Question 2 continued**

(b) Explain why Mary cannot use the data in **Figure 4** to investigate whether the **world** total fertility rate changed between 2000–2005 and 2010–2015.

**(1)**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Question 2 continued**

For the data in **Figure 4**, a country's **total fertility rate** is defined as:

*"The expected number of children a woman who survives to the end of the reproductive age span will have during her lifetime".*

In order to calculate a country's total fertility rate, the following information is required

- The number of women at each age
- The age of the mother for each birth

(c) Suggest a **practical** difficulty that might arise when **collecting** this information.

You should also include a comment on any bias that may arise due to this practical difficulty.

**(3)**

**(Total for Question 2 is 10 marks)**

**3** Mary uses UN data to find out about birth weights of male babies. For two of the countries in her sample, she finds the mean and standard deviation of male baby birth weights, in grams, to the nearest gram.

| Country | Mean | Standard deviation |
|---|---|---|
| Guatemala | 3151 | 551 |
| Austria | 3371 | 596 |

(Data source: http://www.data.un.org)

Assuming that the birth weight of male babies in each country follows a **normal** distribution, and that each birth is independent of all others, calculate the probability that

(i) a randomly selected male baby born in Guatemala has a birth weight of less than 4000 g,

**(2)**

(ii) the total weight of two randomly selected male babies born in **Austria** is less than 6500 g,

**(4)**

**Question 3 continued**

(iii) a randomly selected male baby born in Austria is more than 20% heavier than a randomly selected male baby born in Guatemala.

**(5)**

**(Total for Question 3 is 11 marks)**

**4** Sal manages a large company based in the UK. He reads a report about flexible working. It says that 32% of employees in the UK feel that flexible working is discouraged by their manager and that this proportion is high in comparison to other European countries.

(Data source: https://www.cipd.co.uk/knowledge/work/trends/flexible-working)

Sal's company's policies encourage flexible working. He therefore believes that a smaller proportion of his employees should feel that flexible working is discouraged by their manager than the proportion for the UK as a whole.

Sal asks an independent organisation to carry out a confidential survey about attitudes to flexible working on a random sample of his company's employees.

Of the 200 employees surveyed, 41 said they felt that flexible working is discouraged by their manager.

(a) **Use a suitable distributional approximation** and hypothesis test to investigate whether Sal's belief is correct.

You may assume that employees' responses were independent of one another.

**(7)**

**Question 4 continued**

(b) Explain why it was reasonable to use a distributional approximation in part (a).

Use numerical justification.

**(1)**

(c) Explain why it is unlikely that Sal would have to use a distributional approximation to carry out the test in part (a).

**(1)**

**(Total for Question 4 is 9 marks)**

**5** Wollemia is an endangered Australian tree. Catherine is studying the effect of light level on the chlorophyll concentration in the leaves of wollemia seedlings.

Catherine planted six wollemia seedlings in her laboratory. She ensured that the seedlings were all of the same age and received the same amounts of water and food during the experiment. Catherine planted half of the seedlings in soil with a pH of 4.5 and the other half in soil with a pH of 6.5

For each soil pH level, one of the seedlings was exposed to a low light level, one to a medium light level and one to a high light level.

Allocations to growing conditions were made at random.

After 24 months, Catherine took pieces of leaf from seedlings subjected to each combination of light level and soil pH. She measured the chlorophyll concentration in the leaves in micrograms per square centimetre ($\mu g/cm^2$).

Catherine's results are shown in **Figure 5**.

| | | **Light level** | | | |
|---|---|---|---|---|---|
| | | **Low** | **Medium** | **High** | **Row sums** |
| **Soil pH** | **4.5** | 40.5 | 40.2 | 49.5 | **130.2** |
| | **6.5** | 25.9 | 26.3 | 32.2 | **84.4** |
| | **Column sums** | **66.4** | **66.5** | **81.7** | **214.6** |

**Figure 5**

(Data source: https://dx.doi.org/10.1093/aobpla/plu011)

(a) Name the type of experimental design used by Catherine.

**(1)**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Question 5 continued**

(b) Use Catherine's data to complete the ANOVA table given in **Figure 6**.

| Source | SS | df | MS | F |
|---|---|---|---|---|
| **Soil pH** | | | | |
| **Light level** | | | | |
| **Error** | | | | |
| **Total** | 430.35 | | | |

**Figure 6**

**(6)**

**Question 5 continued**

A **higher** level of chlorophyll in the leaves implies a healthier seedling.

(c) Use your results from part (b) and the information in **Figure 5** to draw a conclusion about any link between **light level** and the **health** of wollemia seedlings.

**(3)**

(d) Did including the factor of soil pH in her experiment provide Catherine with useful additional information?

Explain your answer.

**(2)**

**Question 5 continued**

(e) Explain why Catherine carried out her experiment in the laboratory and made sure that the wollemia seedlings received the same amounts of food and water.

**(2)**

(f) Explain fully, in context, how Catherine's decision to carry out her experiment in the laboratory and to control the growing conditions might limit any generalising she may make in her conclusions.

**(2)**

**(Total for Question 5 is 16 marks)**

**6** (a) State the definition of a statistic.

**(2)**

A biscuit manufacturer claims that each packet contains, on average, 200 g of its biscuits. Lionel wants to test this claim.

He takes a random sample of 12 packets of the biscuits and weighs their contents. The mean weight of his sample is 192 g with a standard error of 4.91 g.

(b) In **Figure 7**, put a tick against the numbers, taken from the above information, that are statistics.

**(1)**

| | |
|---|---|
| 200 g | |
| 192 g | |
| 4.91 g | |

**Figure 7**

(c) Use Lionel's data to calculate a 95% confidence interval for the mean weight of a packet of biscuits.

**(3)**

**Question 6 continued**

(d) Comment on the manufacturer's claim that the average weight of biscuits in a packet is 200 g with reference to your result in (c).

**(2)**

**(Total for Question 6 is 8 marks)**

**7** The Northern Ireland Multiple Deprivation Measure (NIMDM) is used by the Northern Ireland Government to compare levels of social deprivation in different geographical areas of Northern Ireland.

The NIMDM combines ratings of several factors affecting deprivation including Education, Income and Employment.

Eoin is a sociologist. He is comparing rural (countryside) and urban (city) areas of Northern Ireland. He is interested in the associations between Educational deprivation and other kinds of deprivation.

He calculates the measures in **Figure 8** from the NIMDM ratings for 2017.

**A lower NIMDM value indicates greater deprivation.**

| | | **Rural** | **Urban** |
|---|---|---|---|
| **Median** measure of deprivation for Education | | 554 | 328.5 |
| **IQR** of the measure of deprivation for Education | | 252 | 469.25 |
| **Spearman's rank correlation coefficient** (SRCC) between the measures of deprivation for | Education and Income | 0.490 | 0.745 |
| | Education and Employment | 0.779 | 0.863 |

(Data source: https://www.nisra.gov.uk)

**Figure 8**

Using the information in **Figure 8**, write a short report comparing levels of deprivation in rural and urban areas of Northern Ireland, **with a focus on Educational deprivation**.

You are expected to do the following.

- Write **three** distinct comments that each make a **comparison** between rural and urban areas.
- Include comments relating to all three types of measure in **Figure 8** (median, IQR and SRCC).
- Write your comments in context.
- Draw a conclusion that summarises your findings.
- Write the report in a style appropriate for Eoin to read out in a public meeting.

**(6)**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Question 7 continued**

**(Total for Question 7 is 6 marks)**

**TOTAL FOR PAPER IS 80 MARKS**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**BLANK PAGE**

**BLANK PAGE**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**BLANK PAGE**