Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number	Candidate Number

Pearson Edexcel Level 3 GCE

Monday 12th June 2023

Morning (Time: 2 hours)

Paper reference **9ST0/02**

Statistics

Advanced

PAPER 2: Statistical inference

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.

Turn over ▶

R72910A

N:1/1/1/1/1/

Pearson

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

1 100 professional players qualified to take part in the World Cup for a video game.

In each round, players were awarded points based on their position.

The overall positions in the tournament were determined by the total number of points that the players scored from all the rounds. The winning player gained the highest number of final points.

Keito believes that professional players from countries in Asia are better players than those from other continents.

To test this belief, he obtained a sample of 20 players from those in the World Cup. The final points of the players are shown in **Figure 1**.

Final points of players from countries in Asia	Final points of players from countries not in Asia
32	20
26	17
23	15
19	13
16	12
14	11
4	10
3	6
	5
	2
	1
	0

Figure 1

(a) Carry out a non-parametric test to determine whether there is evidence to support Keito's belief.

(9)

Question 1 continued

(b) Suggest **one** improvement that could be made to the test in (a).

(1)

Question 1 continued

Tim suggested using a test on the mean number of points scored by players, from countries in Asia and those not in Asia, using the ***t*-distribution** to investigate Keito's belief.

(c) Give **one** reason why a non-parametric test was chosen in (a), rather than the test suggested by Tim.

(1)

After carrying out his non-parametric test, Keito wrote an article on his blog. An extract from his blog is given in **Figure 2**.

> As my test shows the players from Asian countries in the World Cup did better than those from countries not in Asia, we can conclude that Asian players are better at video games than non-Asian players.
>
> Furthermore it is clear this is due to the living arrangements used by Asian teams to form a team house. The team house culture gives Asian players a huge advantage and is one that the rest of the world has not yet caught up with.

Figure 2

(d) Give **two** criticisms of the comments that Keito has made in his blog.

(2)

(Total for Question 1 is 13 marks)

BLANK PAGE

2 It is believed that colour preference in children is influenced by many factors.

Pippa is a child psychologist who is interested in investigating whether home location affects colour preference.

Colour preferences were tested in a sample of females aged 4-11 years in four different home locations:

- Shipibo villages in the Peruvian Amazon,
- Kastom villages in the highlands of Tanna Island, Vanuatu in the Pacific Islands,
- BaYaka villages in the northern Republic of Congo,
- Brisbane, a large city in Australia.

The results of this experiment are summarised in **Figure 3**.

Observed Frequencies		**Colour Preference**						
		Light Pink	**Pink**	**Red**	**Purple**	**Blue**	**Light Blue**	**Total**
Home Location	**Shipibo**	8	11	9	6	14	7	55
	Kastom	6	9	14	15	11	9	64
	BaYaka	4	8	13	5	14	7	51
	Brisbane	15	32	17	18	11	5	98
	Total	33	60	53	44	50	28	268

[Data Source: https://srcd.onlinelibrary.wiley.com/doi/full/10.1111/cdev.13528]

Figure 3

Some of the expected values for the χ^2 test for an association are shown in **Figure 4**.

Expected Frequencies		**Colour Preference**					
		Light Pink	**Pink**	**Red**	**Purple**	**Blue**	**Light Blue**
Home Location	**Shipibo**		12.31	10.88	9.03	10.26	5.75
	Kastom		14.33	12.66	10.51	11.94	6.69
	BaYaka		11.42	10.09	8.37	9.51	5.33
	Brisbane		21.94	19.38	16.09	18.28	10.24

Figure 4

Question 2 continued

(a) Complete the table in **Figure 4**.

(2)

Question 2 continued

Some of the contributions to the χ^2 test statistic are given in **Figure 5.**

		Colour Preference					
		Light Pink	**Pink**	**Red**	**Purple**	**Blue**	**Light Blue**
Home Location	**Shipibo**	0.223	0.140	0.324		1.362	0.274
	Kastom	0.449	1.981	0.143		0.074	0.800
	BaYaka	0.828	1.023	0.842		2.114	0.524
	Brisbane	0.713	4.612	0.292		2.902	2.680

Figure 5

(b) Test whether colour preference of females aged 4-11 years is independent of home location.

(6)

Question 2 continued

(c) Describe the nature of any association found in (b)

You should include numerical justification.

(2)

It was later discovered that the four samples of children did not have the same proportions of each of the ages from 4 years to 11 years.

(d) Explain why the conclusion in (b) is now likely **not** to be valid and suggest a possible modification to Pippa's experimental design.

(2)

(Total for Question 2 is 12 marks)

3 On the 25th June 2019 the BBC wrote an article comparing key statistics for the highest leagues, Premier and Super, of men's and women's football in England.

Within this article it was stated that there were more goals scored per match in the women's games than in the men's games.

[Source: https://www.bbc.co.uk/news/uk-48742850]

Paul is interested in investigating whether the claim made in 2019 is still true. He found that the mean number of goals scored per match for the 2021/22 season in the men's Premier League was 2.82

He then gathered data on numbers of goals scored for a random sample of the women's Super League matches for the 2021/22 season. The results are shown below.

1 2 3 5 7 4 0 4 5 3 1 0

[Source: https://www.soccerstats.com/table.asp?league=england17&tid=d]

(a) Making any necessary assumptions, perform a hypothesis test to investigate whether, for the 2021/22 season, there were more goals scored per match, on average, in the women's Super League than in the men's Premier League.

(6)

Question 3 continued

A separate claim, made in the same article, is that the women break the rules in football **less** often than men.

A yellow card is given to a player as a warning about involvement in a rule breaking incident.

During the men's Premier League's 2021/22 season, the mean number of yellow cards per player was found to be 3.37

Paul gathered data for the number of yellow cards given to each of a random sample of 100 female players playing in the women's Super League for the same season. Using his data, Paul found that the mean number of yellow cards per player was 1.93 and the **variance** was 2.76

(b) Perform a test to investigate whether women in the Super League, on average, are given **fewer** yellow cards than men in the Premier League.

(5)

(c) Explain why there was no need to make any assumptions with regards to the population distribution when carrying out the test in (b)

(2)

Question 3 continued

(d) Write a short article suitable for the general public comparing the highest leagues of men's and women's football in England.

You should ensure that your answer includes

- references to your conclusions in (a) and (b)
- relevant comments with regards to the original sample data.

(4)

(Total for Question 3 is 17 marks)

BLANK PAGE

4 Alder is a first year psychology student at a university. He decides to use a non-parametric test on some data he has gathered.

He reads some statistics books and decides that either a sign test or the Wilcoxon signed-rank test could be appropriate.

Compare these two tests for Alder.

Your comparison should include comments on

- their similarities,
- their differences,
- the assumptions necessary, if any, for each test, for that test to be valid.

(5)

Question 4 continued

(Total for Question 4 is 5 marks)

5 In March 2011, there was an accident at the Fukushima Nuclear Power Plant in Japan that caused a massive release of radioactive materials into the environment.

An investigation was carried out to see whether the release of radioactive materials affected the *Zizeeria maha*, a common butterfly in Japan.

The **first** generation of butterflies from the Fukushima area was observed in May 2011.

The offspring of the first generation, called the **second** generation, were then observed.

Also, the offspring of the second generation, called the **third** generation, were observed.

For each of these generations, the wing size, cm, for a random sample of male butterflies was measured.

The recorded measurements are shown in **Figure 6**.

	First Generation	Second Generation	Third Generation
	1.43	1.30	1.27
	1.44	1.33	1.31
	1.44	1.38	1.38
	1.43	1.34	1.31
	1.42	1.37	1.35
	1.40	1.33	1.32
	1.45	1.40	1.41
Total	10.01	9.45	9.35

[Data Source https://www.nature.com/articles/srep00570]

Figure 6

The data produced the following summary statistic.

$$\sum x_i^2 = 39.5831$$

(a) Carry out a hypothesis test to investigate whether there is any difference in the average wing size for male butterflies between the three generations.

NOTE: When using Table 7 in the Statistical formulae and tables booklet to obtain the critical value, if the exact critical value is not available you should use the closest critical values either side of the degrees of freedom required.

(10)

Question 5 continued

Question 5 continued

It was later discovered in this investigation that, for each generation, a male butterfly was observed in each of seven different locations surrounding the Fukushima area.

(b) Explain how your analysis in (a) could be adapted using this further information and state the name of the new experimental design.

(2)

(c) Comment on how the experimental error will be affected by using the design stated in (b).

(1)

Question 5 continued

For each of the seven locations, the distance, in km, from the Fukushima Nuclear Power Plant was recorded alongside the average abnormality rate, %, for the male butterflies in the **first** generation.

One of the researchers, Joji, believed that the abnormality rate would be lower for locations further away from the Fukushima Nuclear Power Plant.

The data for the seven locations is shown in **Figure 7**.

Location	A	B	C	D	E	F	G
Distance from Power Plant	20	30	57	58	77	81	128
Abnormality Rate	6.8	6.6	4.1	3.8	1.5	3.8	0.7

Figure 7

(d) Calculate Pearson's product moment correlation coefficient between the distance from the Power Plant and the abnormality rate.

(1)

(e) Making any necessary assumptions, conduct a hypothesis test to investigate Joji's belief.

(4)

(Total for Question 5 is 18 marks)

6 A study investigated the effects of smartphone displays with blue light on the sleepiness of adults playing smartphone games.

[Source: Effects of smartphone use with and without blue light at night in healthy adults: A randomized, double-blind, cross-over, placebo controlled comparison - ScienceDirect]

Two groups of adults took part in the trial.

- One group played games **with** blue light.
- One group played games **without** blue light.

Sleepiness levels were measured for each adult after playing the games.

A *t*-test was carried out to investigate whether the adults playing games **with** blue light had a **decreased** average sleepiness level compared to those playing games **without** blue light.

The test gave a *p*-value of 0.04 and a Cohen's *d* value of 0.49

Use both the *p*-value and the value of Cohen's *d* for this test to make conclusions about the difference between average sleepiness levels for adults playing games with blue light and those playing without blue light.

You may assume that sleepiness levels for all adults are normally distributed.

(4)

Question 6 continued

(Total for Question 6 is 4 marks)

7 Justin investigated how advertising food and drinks to children affects their food consumption.

A group of 20 children was shown an episode of a particular TV show on two consecutive Wednesdays.

One Wednesday the children were shown an episode with adverts for **toys** during the break.

The other Wednesday the children were shown an episode with adverts for **food** during the break.

After they had seen each episode, the children were given a tray of food. The total amount of food each child ate was measured.

(a) Explain which data Justin should use if he wanted to carry out a paired *t*-test as part of his investigation.

(2)

Kareema says a paired *t*-test is **only** valid if both samples are selected from normally distributed populations.

(b) Explain why it is valid to use a paired *t*-test when both samples are from normally distributed populations.

(1)

(c) Explain whether or not it is necessary for **both** samples to be selected from normally distributed populations for a paired *t*-test to be valid.

(2)

Question 7 continued

Justin wanted to make sure that the only factor causing the difference in food eaten was the type of advert the children were shown.

He was concerned that the hunger level of the children may have been different on the two Wednesdays.

Therefore, he also asked the 20 children to rate their hunger on a scale of 1, not hungry, to 5, very hungry, before they watched the episode.

The mean difference of the hunger rates for the children between the two Wednesdays was 0.03 and the standard deviation of these differences was 0.89

[Source: https://livrepository.liverpool.ac.uk/3027509/1/201153268_Oct2018.pdf.pdf]

Justin decided not to use hunger as a factor in his experiment due to this data.

(d) Carry out a hypothesis test to explain Justin's decision.

(5)

Question 7 continued

It is also possible to conduct a paired test with a sign test or the Wilcoxon signed-rank test. These tests are non-parametric.

(e) Explain what is meant by a non-parametric test.

(1)

(Total for Question 7 is 11 marks)

(TOTAL FOR PAPER IS 80 MARKS)