Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number	Candidate Number

Pearson Edexcel Level 3 GCE

Time 2 hours

Paper reference **9ST0/03**

Statistics

Advanced
PAPER 3: Statistics in Practice

You must have:	Total Marks
Statistical formulae and tables booklet Calculator	

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 6 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 ▶

R69434RA

Q:1/1/1/1/1/1/

Pearson

BLANK PAGE

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

1 Following an illness, patients might need therapy in order to help improve their balance. The therapy offered is called standard balance training.

Giovanni, a physiotherapist, thinks that his patients may be helped further by using a video game designed to improve balance.
He designs an experiment to test his theory.

He randomly assigns patients to one of two groups.
One group receives a course of therapy including standard balance training and the video game. The other group receives only standard balance training.

(a) State which of the two groups described above is the **control** group.

(1)

(b) Explain, in context, why Giovanni **randomly** assigns each patient to one of the two groups.

(1)

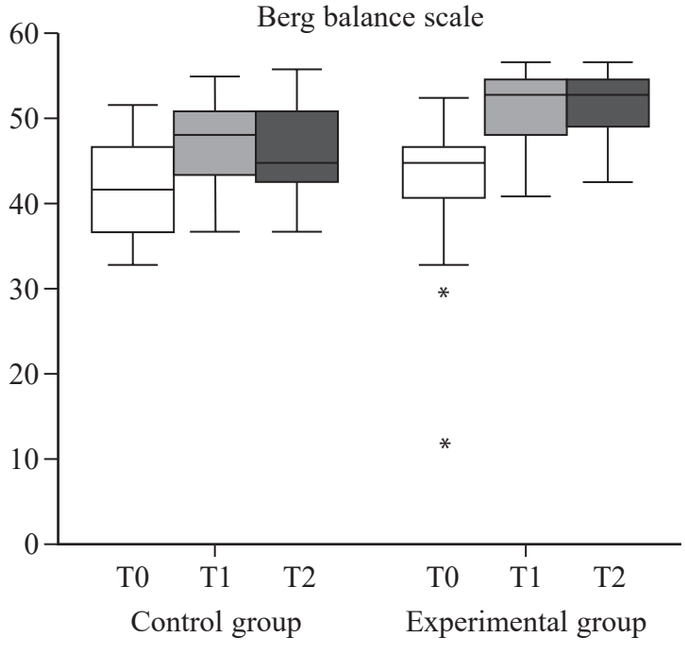
(c) Explain why Giovanni's experiment design is not double-blind.

(1)

Question 1 continued

At the start of the therapy (T0), Giovanni measures the patients' ability to balance using the "Berg balance scale".
He repeats this measure at the end of the course of therapy (T1), and at a follow-up appointment one month later (T2).

Giovanni displays his results using box and whisker plots, shown in **Figure 1**.



[Source: http://dx.doi.org/10.1155/2014/580861]

Figure 1

A **higher score** on the Berg balance scale indicates **better** balance.

(d) Make **four** distinct comments, in context, about Giovanni's experiment and on his results using the information shown in **Figure 1 only**.

(4)

Question 1 continued

(e) Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results.

(2)

(Total for Question 1 is 9 marks)

2 Shelly works as a project manager for a clothing design company. She is in charge of a project to produce a new jacket design. A pattern is then made of her design.

Shelly sends the pattern to the sample production department to be made into a sample jacket. From past experience, it is known that it should take between 5 and 13 working hours for a sample jacket to be produced from a pattern.

The uniform distribution is considered a suitable model for the time taken to produce such a sample jacket from a pattern.

A diagram of Shelly's modelling distribution is shown in **Figure 2**.

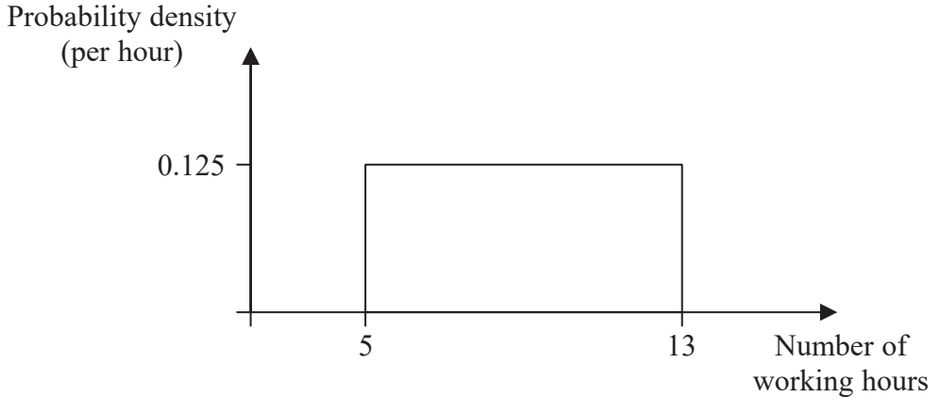


Figure 2

(a) Explain why the height of the modelling distribution in **Figure 2** is 0.125

(1)

Question 2 continued

Use Shelly's model and **Figure 2** to answer parts (b) to (e).

(b) State the expected time that it will take to produce her sample jacket from the pattern.

(1)

(c) State the probability that it will take less than 4 hours to produce her sample jacket from the pattern.

(1)

Question 2 continued

One week, Shelly sends the pattern to the sample production department on Tuesday, 7 hours before it closes for the day.

(d) Calculate the probability that her sample jacket will be produced from the pattern before the sample production department closes on Tuesday.

(1)

That same week, Shelly's boss would like the sample jacket to be produced from the pattern by 11 am on Wednesday.
The sample production department opens at 8 am.

(e) Given that the sample jacket was **not** produced from the pattern before the sample production department closed on Tuesday, calculate the probability that it is ready by 11 am on Wednesday.

(2)

(Total for Question 2 is 6 marks)

3 Klazine is a researcher interested in nutrition. She believes that involving a child in preparing their own meal affects what they choose to eat at that meal.

She conducted a study involving 47 children aged 6 to 10 years.

25 children ate a meal that they had prepared with their parent.

The other 22 children ate a meal that their parent had prepared alone.

All of the meals contained the same four ingredients: pasta, chicken, cauliflower, and salad.

Klazine measured the total weight, in grams (g), of each ingredient eaten by the children in each group. Her results for **salad** are shown in **Figure 3**.

		Salad eaten (g)	
	n	**Mean**	**SD**
Child prepared with parent	25	96.4	61.5
Parent prepared alone	22	54.8	35.4

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 3

(a) Calculate a 95% confidence interval for the weight of salad eaten by the children whose parent prepared the meal alone.

(3)

Question 3 continued

A 95% confidence interval for the weight, in grams, of salad eaten by the children who prepared the meal with their parent is (71.0, 121.8)

(b) Give a full explanation of the evidence provided about Klazine's belief by the confidence interval in (a), together with the interval given above.

(3)

Question 3 continued

Klazine measured the weight of cauliflower, in grams (g), eaten by the children in each group. Her results are shown in **Figure 4**.

		Cauliflower eaten (g)	
	n	**Mean**	**SD**
Child prepared with parent	25	110.5	50.1
Parent prepared alone	22	89.7	51.3

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 4

(c) Making any necessary assumptions, use a ***t*-test** to investigate whether the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children whose parent prepared the meal alone.

(8)

Question 3 continued

There is an assumption about variance that is necessary for the *t*-test in (c) to be valid.

(d) State the assumption about variance, in context, and comment on its appropriateness, given the data in **Figure 4**.

(2)

Question 3 continued

Klazine's colleague, Aurora, is concerned that the sample sizes used in their study are too small to be able to detect a difference between the two groups of children.

(e) Discuss Aurora's concern about **sample sizes**.

Include **one** reference to the statistical techniques applied in (a) and in (c) and their conclusions and **one** reference to the effect of increasing sample sizes.

(2)

(f) If Klazine uses a much larger sample of children in her study, what effect would you expect this to have on the confidence intervals in (a) and in (b)?

(1)

(Total for Question 3 is 19 marks)

4 Petra is an engineer who works on an offshore wind farm.

Petra organises the repairs of wind turbines that fail. The wind farm contract states that wind turbines must be repaired on the same day that they fail.

Failures appear to happen at random, and at a fairly consistent rate over the year.

In order to organise repairs, Petra uses a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

[Data source: https://doi.org/10.1177%2F0957650915597560]

Use Petra's model to answer (a) to (f).

(a) Calculate the standard deviation of the number of failures per year for a wind turbine at the wind farm.

(1)

(b) Find the probability that a wind turbine at the wind farm fails exactly three times in a given year.

(1)

(c) Find the probability that a wind turbine at the wind farm fails more than seven times over a three-year period.

(2)

Question 4 continued

(d) State fully the distribution of the time, in years, between successive failures of a wind turbine at the wind farm.

(2)

(e) Calculate the expected time, in years, between successive failures of a wind turbine at the wind farm.

(1)

(f) Given that a particular wind turbine at the wind farm has not failed for two consecutive years, calculate the probability that it fails within the next six months.

(3)

Question 4 continued

The wind farm has six wind turbines. During the past five years, there have been 84 wind turbine failures at the wind farm.

(g) Show how Petra used this information to calculate the value of λ

(1)

(h) Make **three** criticisms, in context, of Petra's use of a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

You may consider both the use of a Poisson model and the value of λ

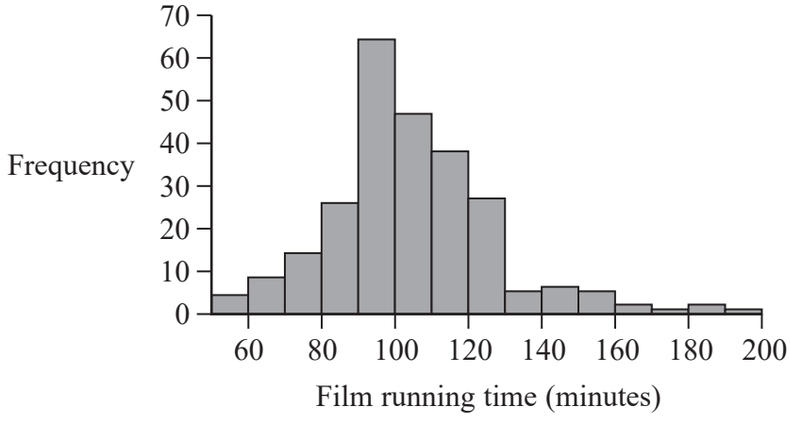
(3)

(Total for Question 4 is 14 marks)

5 Dennis and Viola are doing a course in film studies. They want to conduct a study into the factors affecting film running times. They start by investigating the distribution of film running times.

Dennis and Viola extract a random sample of 250 films from a database containing the names and dates of all films ever made and look up the running time of each film in their sample on the internet.

They use graphing software to produce a histogram of their results, shown in **Figure 5**.



[Data sources: www.cinemetrics.lv, www.imdb.com]

Figure 5

Viola believes that a normal distribution might be a suitable model for film running times.

(a) Explain **one** feature of **Figure 5** that supports Viola's belief.

(1)

(b) Explain **one** feature of **Figure 5** that does not support Viola's belief.

(1)

Question 5 continued

Dennis and Viola agree that they should perform a goodness-of-fit test to investigate whether a normal distribution would be a suitable model for film running times.

For their sample of 250 films, the mean running time is 105.26 minutes and the standard deviation of the running times is 22.18 minutes.

Part of their working is shown in **Figure 6**.

Film running time (minutes)	Observed frequency	Probability (4 d.p.)	Expected frequency (2 d.p.)
$x \leqslant 60$	4	0.0206	5.16
$60 < x \leqslant 80$	22	0.1067	26.68
$80 < x \leqslant 100$	90	s	u
$100 < x \leqslant 120$	85	t	v
$120 < x \leqslant 140$	32	0.1945	48.63
$140 < x \leqslant 160$	11	0.0518	12.96
$x > 160$	6	0.0068	1.70
Total	250	1	250

Figure 6

(c) Find the values of s, t, u and v missing from **Figure 6**.

(4)

Question 5 continued

(d) Explain, with a reason, what adjustment Viola and Dennis must make to the data in **Figure 6** before they can carry out the χ^2 goodness-of-fit test.

(1)

Question 5 continued

Viola and Dennis make the necessary adjustment to their data.

They calculate the χ^2 test statistic to be 13.04 (to 2 d.p.)

(e) Use this information to complete their χ^2 goodness-of-fit test.

You must state appropriate hypotheses.

(4)

(Total for Question 5 is 11 marks)

6 Russell and Robert are interested in sport. They believe that wearing red may affect the success of a person or team in a sporting competition.

Russell finds the following information about the 2004 Olympic games.

- Male combatants in combat sports were randomly assigned either a red or a blue outfit.
- There was a total of 21 rounds of fighting for these combat sports.
- One round had an equal number of red and blue winners.
- Four rounds had **more blue** than red winners
- Sixteen rounds had **more red** than blue winners.

[Source: doi:10.1038/435293a]

(a) Making any necessary assumptions, use Russell's data to investigate their belief.

Make a full conclusion, in context, about any effect seen, or otherwise.

(7)

Question 6 continued

Question 6 continued

Robert collected data about the performance of five teams that sometimes wore a red kit at the Euro 2004 men's football competition.

He awarded one point for a win, zero for a draw and minus one point for a loss (disregarding penalty shoot-outs). He then found the average points scored when each team wore red and when they did not wear red.

Opposing football teams never wear the same colour kit as each other.

Robert's results are shown in **Figure 7**.

	Total points		Number of games		Average points	
Team	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**
Croatia	0	–1	1	2	0	–0.5
Czechia	1	2	1	4	1	0.5
England	1	0	1	3	1	0
Latvia	–1	–1	2	1	–0.5	–1
Spain	1	–1	2	1	0.5	–1

[Data source: doi.org/10.1038/435293a]

Figure 7

(b) Making any necessary assumptions, use the **average points** data in **Figure 7** and a **paired t-test** to investigate Russell and Robert's belief that wearing red may affect the success of a team.

Make a full conclusion, in context, about any effect seen, or otherwise.

(8)

Question 6 continued

Question 6 continued

(c) Considering the data and tests used in (a) and (b), make **four** distinct criticisms.

In your answer you might consider

- the size of sample provided,
- the sampling design,
- the test used and its assumptions.

(4)

Question 6 continued

Charlotte reads Russell and Robert's research and tells her women's netball team that they should change to wearing red to improve their results.

(d) Make **two** comments about Charlotte's statement in light of the data given and the conclusions to the tests in (a) and (b).

(2)

(Total for Question 6 is 21 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

BLANK PAGE