Write your name here

Surname	Other names

Pearson
Edexcel GCE

Centre Number	Candidate Number

Statistics S4

Advanced/Advanced Subsidiary

Tuesday 26 June 2018 – Morning **Time: 1 hour 30 minutes**	Paper Reference **6686/01**

You must have: Mathematical Formulae and Statistical Tables (Pink)	Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or 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). Coloured pencils and highlighter pens must not be used.
- **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.
- Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- The total mark for this paper is 75.
- 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.

Turn over ▶

P51570A

1/1/1/1/

P 5 1 5 7 0 A 0 1 2 4

Pearson

Leave blank

1. A machine fills packets with almonds. The weight, in grams, of almonds in a packet is modelled by $N(\mu, \sigma^2)$. To check that the machine is working properly, a random sample of 10 packets is selected and unbiased estimates for μ and σ^2 are

$$\bar{x} = 202 \quad \text{and} \quad s^2 = 3.6$$

Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean weight of almonds in a packet is more than 200 g.

(5)

Leave blank

Question 1 continued

Q1

(Total 5 marks)

Leave blank

2. Jeremiah currently uses a *Fruity* model of juicer. He agrees to trial a new model of juicer, *Zesty*. The amounts of juice extracted, x ml, from each of 9 randomly selected oranges, using the *Zesty* are summarised as

$$\sum x = 468 \qquad \sum x^2 = 24560$$

Given that the amounts of juice extracted follow a normal distribution,

(a) calculate a 95% confidence interval for

(i) the mean amount of juice extracted from an orange using the *Zesty*,

(ii) the standard deviation of the amount of juice extracted from an orange using the *Zesty*.

(9)

Jeremiah knows that, for his *Fruity*, the mean amount of juice extracted from an orange is 38 ml and the standard deviation of juice extracted from an orange is 5 ml.

He decides that he will replace his *Fruity* with a *Zesty* if both

- the mean for the *Zesty* is more than 20% higher than the mean for his *Fruity*

  and
- the standard deviation for the *Zesty* is less than 5.5 ml.

(b) Using your answers to part (a), explain whether or not Jeremiah should replace his *Fruity* with the *Zesty*.

(4)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

(Total 13 marks)

Q2

Leave blank

3. A random sample of 8 students is selected from a school database.

Each student's reaction time is measured at the start of the school day and again at the end of the school day. The reaction times, in milliseconds, are recorded below.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Reaction time at the start of the school day	10.8	7.2	8.7	6.8	9.4	10.9	11.1	7.6
Reaction time at the end of the school day	10	6.1	8.8	5.7	8.7	8.1	9.8	6.8

(a) State one assumption that needs to be made in order to carry out a paired *t*-test.

(1)

The random variable R is the reaction time at the start of the school day minus the reaction time at the end of the school day. The mean of R is μ.

John uses a paired *t*-test to test the hypotheses

$$H_0: \mu = m \qquad H_1: \mu \neq m$$

Given that H_0 is rejected at the 5% level of significance but accepted at the 1% level of significance,

(b) find the ranges of possible values for m.

(9)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 10 marks)

Q3

Leave blank

4. A glue supplier claims that *Goglue* is stronger than *Tackfast*. A company is presently using *Tackfast* but agrees to change to *Goglue* if, at the 5% significance level,

- the standard deviation of the force required for *Goglue* to fail is not greater than the standard deviation of the force required for *Tackfast* to fail

  and

- the mean force required for *Goglue* to fail is more than 4 newtons greater than the mean force for *Tackfast* to fail.

A series of trials is carried out, using *Goglue* and *Tackfast*, and the glues are tested to destruction. The force, x newtons, at which each glue fails is recorded.

	Sample size (n)	Sample mean ($\bar{x}$)	Standard deviation (s)
Tackfast (T)	6	5.27	0.31
Goglue (G)	5	10.12	0.66

It can be assumed that the force at which each glue fails is normally distributed.

(a) Test, at the 5% level of significance, whether or not there is evidence that the standard deviation of the force required for *Goglue* to fail is greater than the standard deviation of the force required for the *Tackfast* to fail. State your hypotheses clearly.

(5)

The supplier claims that the mean force required for its *Goglue* to fail is more than 4 newtons greater than the mean force required for *Tackfast* to fail.

(b) Stating your hypotheses clearly and using a 5% level of significance, test the supplier's claim.

(7)

(c) Show that, at the 5% level of significance, the supplier's claim will be accepted if $\bar{X}_G - \bar{X}_T > 4.55$, where $\bar{X}_G$ and $\bar{X}_T$ are the mean forces required for *Goglue* to fail and *Tackfast* to fail respectively.

(2)

Later, it was found that an error had been made when recording the results for *Goglue*. This resulted in all the forces recorded for *Goglue* being 0.5 newtons more than they should have been. The results for *Tackfast* were correct.

(d) Explain whether or not this information affects the decision about which glue the supplier decides to use.

(3)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 17 marks)

Q4

Leave blank

5. A machine makes posts. The length of a post is normally distributed with unknown mean μ and standard deviation 4 cm.

A random sample of size n is taken to test, at the 5% significance level, the hypotheses

$$H_0 : \mu = 150 \qquad H_1 : \mu > 150$$

(a) State the probability of a Type I error for this test.

(1)

The manufacturer requires the probability of a Type II error to be less than 0.1 when the actual value of μ is 152

(b) Calculate the minimum value of n.

(10)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 11 marks)

Q5

Leave blank

6. The continuous random variable X has probability density function $f(x)$

$$f(x) = \begin{cases} \dfrac{x}{2\theta^2} & 0 \leqslant x \leqslant 2\theta \\ 0 & \text{otherwise} \end{cases}$$

where θ is a constant.

(a) Use integration to show that $E(X^N) = \dfrac{2^{N+1}}{N+2}\theta^N$

(3)

(b) Hence

(i) write down an expression for $E(X)$ in terms of θ

(ii) find $\text{Var}(X)$ in terms of θ

(3)

A random sample $X_1, X_2, \ldots, X_n$ where $n \geqslant 2$ is taken to estimate the value of θ

The random variable $S_1 = q\overline{X}$ is an unbiased estimator of θ

(c) Write down the value of q and show that S_1 is a consistent estimator of θ

(3)

The continuous random variable Y is independent of X and is uniformly distributed over the interval $[0, \dfrac{2\theta}{3}]$, where θ is the same unknown constant as in $f(x)$.

The random variable $S_2 = aX + bY$ is an unbiased estimator of θ and is based on one observation of X and one observation of Y.

(d) Find the value of a and the value of b for which S_2 has minimum variance.

(7)

(e) Show that the minimum variance of S_2 is $\dfrac{\theta^2}{11}$

(1)

(f) Explain which of S_1 or S_2 is the better estimator for θ

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 19 marks)

TOTAL FOR PAPER: 75 MARKS

END