1. It is suggested that a Poisson distribution with parameter λ can model the number of currants in a currant bun. A random bun is selected in order to test the hypotheses H_0: $\lambda = 8$ against H_1: $\lambda \neq 8$, using a 10% level of significance.

(a) Find the critical region for this test, such that the probability in each tail is as close as possible to 5%.

(5)

(b) Given that $\lambda = 10$, find

(i) the probability of a type II error,

(ii) the power of the test.

(4)
(Total 9 marks)

2. A doctor believes that the span of a person's dominant hand is greater than that of the weaker hand. To test this theory, the doctor measures the spans of the dominant and weaker hands of a random sample of 8 people. He subtracts the span of the weaker hand from that of the dominant hand. The spans, in mm, are summarised in the table below.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Dominant hand	202	251	215	235	210	195	191	230
Weaker hand	195	249	218	234	211	197	181	225

Test, at the 5% significance level, the doctor's belief.

(Total 9 marks)

3. A grocer receives deliveries of cauliflowers from two different growers, *A* and *B*. The grocer takes random samples of cauliflowers from those supplied by each grower. He measures the weight x, in grams, of each cauliflower. The results are summarised in the table below.

	Sample size	$\sum x$	$\sum x^2$
A	11	6600	3960540
B	13	9815	7410579

(a) Show, at the 10% significance level, that the variances of the populations from which the samples are drawn can be assumed to be equal by testing the hypothesis $H_0: \sigma_A^2 = \sigma_B^2$ against hypothesis $H_1: \sigma_A^2 \neq \sigma_B^2$.

(You may assume that the two samples come from normal populations.)

(6)

The grocer believes that the mean weight of cauliflowers provided by *B* is at least 150 g more than the mean weight of cauliflowers provided by *A*.

(b) Use a 5% significance level to test the grocer's belief.

(8)

(c) Justify your choice of test.

(2)
(Total 16 marks)

4. A beach is divided into two areas *A* and *B*. A random sample of pebbles is taken from each of the two areas and the length of each pebble is measured. A sample of size 26 is taken from area *A* and the unbiased estimate for the population variance is $s_A^2 = 0.495$ mm^2. A sample of size 25 is taken from area *B* and the unbiased estimate for the population variance is $s_B^2 = 1.04$ mm^2.

(a) Stating your hypotheses clearly test, at the 10% significance level, whether or not there is a difference in variability of pebble length between area *A* and area *B*.

(5)

(b) State the assumption you have made about the populations of pebble lengths in order to carry out the test.

(1)
(Total 6 marks)

5. A random sample of 10 mustard plants had the following heights, in mm, after 4 days growth.

5.0, 4.5, 4.8, 5.2, 4.3, 5.1, 5.2, 4.9, 5.1, 5.0

Those grown previously had a mean height of 5.1 mm after 4 days. Using a 2.5% significance level, test whether or not the mean height of these plants is less than that of those grown previously.

(You may assume that the height of mustard plants after 4 days follows a normal distribution.)

(Total 9 marks)

6. A train company claims that the probability p of one of its trains arriving late is 10%. A regular traveller on the company's trains believes that the probability is greater than 10% and decides to test this by randomly selecting 12 trains and recording the number X of trains that were late. The traveller sets up the hypotheses H_0: $p = 0.1$ and H_1: $p > 0.1$ and accepts the null hypothesis if $x \leq 2$.

(a) Find the size of the test.

(1)

(b) Show that the power function of the test is

$$1 - (1 - p)^{10} (1 + 10p + 55p^2).$$

(4)

(c) Calculate the power of the test when

(i) $p = 0.2$,

(ii) $p = 0.6$.

(3)

(d) Comment on your results from part (c).

(1)
(Total 9 marks)

7. (a) Define

(i) a Type I error,

(ii) a Type II error.

(2)

A small aviary, that leaves the eggs with the parent birds, rears chicks at an average rate of 5 per year. In order to increase the number of chicks reared per year it is decided to remove the eggs from the aviary as soon as they are laid and put them in an incubator. At the end of the first year of using an incubator 7 chicks had been successfully reared.

(b) Assuming that the number of chicks reared per year follows a Poisson distribution test, at the 5% significance level, whether or not there is evidence of an increase in the number of chicks reared per year. State your hypotheses clearly.

(4)

(c) Calculate the probability of the Type I error for this test.

(3)

(d) Given that the true average number of chicks reared per year when the eggs are hatched in an incubator is 8, calculate the probability of a Type II error.

(2)
(Total 11 marks)

8. Two methods of extracting juice from an orange are to be compared. Eight oranges are halved. One half of each orange is chosen at random and allocated to Method *A* and the other half is allocated to Method *B*. The amounts of juice extracted, in ml, are given in the table.

	Orange							
	1	2	3	4	5	6	7	8
Method *A*	29	30	26	25	26	22	23	28
Method *B*	27	25	28	24	23	26	22	25

One statistician suggests performing a two-sample *t*-test to investigate whether or not there is a difference between the mean amounts of juice extracted by the two methods.

(a) Stating your hypotheses clearly and using a 5% significance level, carry out this test.

(You may assume $\bar{x}_A = 26.125$, $s_A^2 = 7.84$, $\bar{x}_B = 25$, $s_B^2 = 4$ and $\sigma_A^2 = \sigma_B^2$)

(7)

Another statistician suggests analysing these data using a paired *t*-test.

(b) Using a 5% significance level, carry out this test.

(9)

(c) State which of these two tests you consider to be more appropriate. Give a reason for your choice.

(1)
(Total 17 marks)

1.	(a)	$P(X \le c_1) \le 0.05$; $P(X \le 3 \mid \lambda = 8) = 0.0424 \Rightarrow X \le 3$ $P(X \le 4 \mid \lambda = 8) = 0.0996 \Rightarrow X \le 3$	M1; A1	
		$P(X \ge c_2) \le 0.05$; $P(X \ge 4 \mid \lambda = 8) = 0.0342 \Rightarrow X \ge 13$ $P(X \ge 13 \mid \lambda = 8) = 0.0638 \Rightarrow X \ge 13$	M1; A1	
		$\therefore$ critical region is $\{X \le 3 \cup X \ge 13\}$	A1 ft	5
	(b)	(i) $P(4 \le X \le 12 \mid \lambda = 10) = P(X \le 12) - P(X \le 3)$	M1 M1	
		$= 0.7916 - 0.0103$ $= \underline{0.7813}$	A1	
		(ii) Power $= 1 - 0.7813 = \underline{0.2187}$	B1 ft	4

[9]

2.	d: 7 2 –3 1 –1 –2 10 5	M1	
	$\Sigma d = 19$; $\Sigma d^2 = 193$		
	$\therefore \bar{d} = \frac{19}{8} = \underline{2.375}$; $S_d^2 = \frac{1}{7}\left\{193 - \frac{19^2}{8}\right\} = \underline{21.125}$	B1; M1 A1	
	$H_0: \mu_D = 0$; $H_1: \mu_D > 0$ *both*	B1	
	$t = \dfrac{2.375 - 0}{\sqrt{\frac{21.125}{8}}} = \underline{1.4615.....}$	M1	
	AWRT 1.46	A1	
	$\upsilon = 7 \Rightarrow$ critical region: $t > 1.895$	B1	
	Since 1.4615… is <u>not</u> in the critical region there is insufficient evidence to reject H_0 and we conclude that there is in sufficient evidence to support the doctors' belief.	A1 ft	

[9]

Alternative:

Use of 2 sample t-test $\Rightarrow$ 6/9 max

$S_p^2 = \dfrac{7 \times 440.125 + 7 \times 501.357}{8 + 8 - 2} = \underline{470.74}$	M1 A1
$t = \dfrac{216.125 - 213.75}{\sqrt{470.74\left(\frac{1}{8} + \frac{1}{8}\right)}} = \underline{0.2189\ldots}$	M1 A1
CR: $t > 1.761$	B1
Conclusion as above	A1 ft

3.	(a)	$S_A^2 = \frac{1}{10}\{3960540 - \frac{6600^2}{11}\} = \underline{54.0}$	B1	
		$S_B^2 = \frac{1}{12}\{7410579 - \frac{9815^2}{13}\} = \underline{21.1\dot{6}}$	B1	
		H_0: $\sigma_A^2 = \sigma_B^2$; H_1: $\sigma_A^2 \neq \sigma_B^2$	B1	
		CR: $F_{10,\,12} > 2.75$		
		$S_A^2 / S_B^2 = \frac{54.0}{21.1\dot{6}} = 2.55118\ldots$	M1 A1	
		Since 2.55118… is not in the critical region, we can assume that the variances of A and B are equal.	B1	6
	(b)	H_0: $\mu_B = \mu_A + 150$; H_1: $\mu_B > \mu_A + 150$ *both*	B1	
		CR: $t_{22}(0.05) > 1.717$	B1	
		$S_p^2 = \frac{10 \times 54.0 + 12 \times 21.1\dot{6}}{22} = \underline{36.\dot{0}\dot{9}}$	M1 A1	
		$t = \frac{1755 - 6001 - 150}{\sqrt{36.0909\ldots\left(\frac{1}{11} + \frac{1}{13}\right)}} = \underline{2.03157}$	M1 A1	
		AWRT 2.03	A1	
		Since 2.03… is in the critical region we reject H_0 and conclude that the mean weight of cauliflowers from *B* exceeds that from *A* by at least 150g.	A1 ft	8
	(c)	Samples from normal populations	B1 B1	2
		Any two sensible verifications		
		Equal variances		
		Independent samples		

[16]

4.	(a)	H_0: $\sigma_A^2 = \sigma_B^2$, H_1: $\sigma_A^2 \neq \sigma_B^2$	both	B1	
		critical values $F_{24.25} = 1.96$ and $\frac{1}{F_{24.25}} = 0.510$	both	B1	
		$\frac{s_B^2}{s_A^2} = 2.10$ or $\frac{s_A^2}{s_B^2} = 0.476$	both	M1 A1	
		Since 2.10 or 0.476 are in the critical region we reject H_0 and conclude there is evidence of a difference in variability of pebble length between area A and B		A1∫	5

	(b)	The populations of pebble lengths are normal.		B1	1	
						[6]
5.		H_0: $\mu = 5.1$, H_1: $\mu < 5.1$	both	B1		
		$v = 9$	9	B1		
		Critical Region $t < -2.262$		B1		
		$\bar{x} = 4.91$	4.91	B1		
		$s^2 = \dfrac{241.89 - 10 \times (4.91)^2}{9} = 0.0899$		M1		
		$s = 0.300$	awrt 0.0899 or 0.300	A1		
		$t = \dfrac{4.91 - 5.1}{\frac{0.3}{\sqrt{10}}} = -2.00$		M1 A1		
		There is no evidence to suggest that the mean height is less than those grown previously	context	A1∫	9	
						[9]
6.	(a)	$1 - 0.8891 = 0.1109$		B1	1	
	(b)	$1 - (P(0) + P(1) + p(2))$		M1		
		$= 1 - ((1 - p)^{12} + 12p(1 - p)^{11} + 66p^2(1 - p)^{10})$		M1 A1		
		$= 1 - (1 - p)^{10}\,((1 - p)^2 + 12p(1 - p) + 66p^2)$				
		$= 1 - (1 - p)^{10}\,(1 + 10p + 55p^2)$ ****given****	cso	A1	4	
	(c)	(i) $1 - 0.5583 = 0.442$		M1 A1		
		$1 - 0.00281 = 0.997$		A1	3	
		(ii) The test is more discriminating for the larger value of p		B1	1	
						[9]
7.	(a)	(i) Type I – H_0 rejected when it is true		B1		
		Type II – H_0 is accepted when it is false		B1	2	
	(b)	H_0: $\lambda = 5$, H_1: $\lambda > 5$	both	B1		
		$P((X \geq 7 \mid \lambda = 5) = 1 - 0.7622 = 0.2378 > 0.05$		M1 A1		
		(**OR** $P(X \geq 9) = 0.0681$, $P(X \geq 10) = 0.0318$, CV = 10, 7 not in CR		M1 A1		
		probabs & 10, dnr				
		No evidence of an increase in the number of chicks reared per year. *Context*		A1	4	

	(c)	$P(X \geq c \mid \lambda = 5) < 0.05$		M1	
		$P(X \geq 9) = 0.0681$, $P(X \geq 10) = 0.0318$, $c = 10$		M1	
		may be seen in (b)			
		P(Type I Error) = 0.0318		A1	3
	(d)	$P(X \leq 9) \mid \lambda = 8) = 0.7166$		M1 A1	
		(OR if $c = 9$ in (d), $P(X \leq 8 \mid \lambda = 8) = 0.5925$		M1 A1	2

[11]

8.	(a)	$s_p^2 = \dfrac{7 \times 7.84 + 7 \times 4}{7 + 7} = 5.92$		M1	
		$s_p = 2.433105$	awrt 2.43	A1	
		$H_0 = \mu_A = \mu_B$, H_1: $\mu_A \neq \mu_B$	both	B1	
		$t = \dfrac{26.125 - 25}{2.43\sqrt{\frac{1}{8} + \frac{1}{8}}} = 0.92474$	awrt 0.925	M1 A1	
		$t_{14}(2.5\%) = 2.145$	2.145	B1	
		Insufficient evidence to reject H_0 that there is no difference in the means.		A1∫	7
	(b)	d = M1 –M2			
		2,5,-2,1,3,-4,1,3		M1	
		$\bar{d} = \dfrac{9}{8} = 1.125$	1.125	B1	
		$s_d^2 = \dfrac{69 - 8 \times 1.125^2}{7} = 8.410714$	awrt 8.41	M1 A1	
		H_0: $\mu_d = 0$, H_1: $\mu_d \neq 0$	both	B1	
		$t = \dfrac{1.125}{\sqrt{\frac{8.41}{8}}} = 1.0972$	awrt 1.10	M1 A1	
		$t_7(2.5\%) = 2.365$	2.365	B1	
		There is no significant evidence of a difference between method A and method B.		A1∫	9
	(c)	Paired sample as they are two measurements on the same orange		B1	1

[17]

1. Use of the Poisson distribution table to find the appropriate critical region was not as well understood as was expected. Too many candidates did not show their working and they then obtained the wrong upper critical value. The definition of a Type II error was not always known but most candidates knew to subtract their probability of a Type II error from one to obtain the power. This is an area of the specification that candidates need to improve upon.

2. The majority of the candidates recognised the need to use a paired *t*-test with only the weaker ones using a test for two means. The hypotheses were not always correctly specified even though the correct test was carried out and some candidates did not give their conclusion relative to the context of the question.

3. Many of the candidates scored full marks in part (a). In part (b) too many of them could not specify the hypotheses correctly since they did not know how to cope with the 150g. Surprisingly many of the candidates did not score both marks in the final part of this question.

4. The question proved to be a friendly starter for the majority, with many candidates gaining full marks. The most common error was in the conclusion where the context was omitted.

5. Candidates had prepared well and solutions given for this question were very good. Hypotheses were stated accurately and the formula was applied well, with clear conclusions that gained full marks.

6. The arithmetic required in parts (a) and (c) was done well and some good solutions were seen for part (b). Some weaker candidates could not produce convincing arguments in part (b) and often missed crucial steps in the working to arrive at the given answer. Conclusions in part (d) interpreted the values well, but many gave trivial statements that gained no credit.

7. The definitions were remembered well for part (a) and the test in part (b) was often correct. However parts (c) and (d) discriminated between candidates, with many incorrect probabilities. Some solutions interpreted the critical region to be above 0.05 and so lost an accuracy mark in part (c), others struggled with the regions required and simply guessed at the probabilities.

8. This was done very well by a large number of candidates. They had clearly prepared well for the examination and it was not unusual to see the majority of the marks being awarded for this question. Occasional accuracy errors were seen, as were some incorrect statements of the hypotheses in part (b) together with missing context in the conclusions. On the whole the responses were very good.