1. Past experience shows that 35% of the senior pupils in a large school know the regulations about bringing cars to school. The head teacher addresses this subject in an assembly, and afterwards a random sample of 120 senior pupils is selected. In this sample it is found that 50 of these pupils know the regulations. Use a suitable approximation to test, at the 10% significance level, whether there is evidence that the proportion of senior pupils who know the regulations has increased. Justify your approximation.

[11]

2. An examination board is developing a new syllabus and wants to know if the question papers are the right length. A random sample of 50 candidates was given a pre-test on a dummy paper. The times, t minutes, taken by these candidates to complete the paper can be summarised by

$$n = 50, \qquad \sum t = 4050, \qquad \sum t^2 = 329\,800.$$

Assume that times are normally distributed.

i. Estimate the proportion of candidates that could not complete the paper within 90 minutes.

[6]

ii. Test, at the 10% significance level, whether the mean time for all candidates to complete this paper is 80 minutes. Use a two-tail test.

[7]

iii. Explain whether the assumption that times are normally distributed is necessary in answering
   a. part **(i)**,
   b. part **(ii)**.

[2]

3. Records for a doctors' surgery over a long period suggest that the time taken for a consultation, T minutes, has a mean of 11.0. Following the introduction of new regulations, a doctor believes that the average time has changed. She finds that, with new regulations, the consultation times for a random sample of 120 patients can be summarised as

$$n = 120, \sum t = 1411.20, \sum t^2 = 18\,737.712.$$

i. Test, at the 10% significance level, whether the doctor's belief is correct.

[11]

ii. Explain whether, in answering part **(i)**, it was necessary to assume that the consultation times were normally distributed.

[1]

4. It is known that the lifetime of a certain species of animal in the wild has mean 13.3 years. A zoologist reads a study of 50 randomly chosen animals of this species that have been kept in zoos. According to the study, for these 50 animals the sample mean lifetime is 12.48 years and the population variance is 12.25 years^2.

i. Test at the 5% significance level whether these results provide evidence that animals of this species that have been kept in zoos have a shorter expected lifetime than those in the wild.

[7]

ii. Subsequently the zoologist discovered that there had been a mistake in the study. The quoted variance of 12.25 years^2 was in fact the sample variance. Determine whether this makes a difference to the conclusion of the test.

[5]

iii. Explain whether the Central Limit Theorem is needed in these tests.

[1]

5. In the past, the time spent in minutes, by customers in a certain library had mean 32.5 and standard deviation 8.2.
Following a change of layout in the library, the mean time spent in the library by a random sample of 50 customers is found to be 34.5 minutes.
Assuming that the standard deviation remains at 8.2, test at the 5% significance level whether the mean time spent by customers in the library has changed. [7]

6. The times taken by employees to complete a task are normally distributed with standard deviation 2.6 minutes. A manager claims that the mean time is 15.5 minutes but an employee suspects that the mean time is greater than this. He intends to carry out a hypothesis test at the 5% significance level to test this claim. He records the times taken by a random sample of 12 employees.

(a) Find the critical region for the test. [3]

(b) The total of the times taken by the 12 employees was 202.1 minutes. Carry out the test. [5]

7. In the past, the time spent by customers in a certain shop had mean 10.5 minutes and standard deviation 4.2 minutes. Following a change of layout in the shop, the mean time spent in the shop by a random sample of 50 customers is found to be 12.0 minutes.

(a) Assuming that the standard deviation is unchanged, test at the 1% significance level whether the mean time spent by customers in the shop has changed. [7]

(b) Another random sample of 50 customers is chosen and a similar test at the 1% significance level is carried out. Given that the population mean time has not changed, state the probability that the conclusion of the test will be that the population mean time **has** changed. [1]

END OF QUESTION paper

Mark scheme

Question			Answer/Indicative content	Marks	Part marks and guidance	
1			H_0: $p = 0.35$ H_1: $p > 0.35$	B2	One error (e.g. μ, no symbol, 2-tailed) B1, but $\overline{x}$, t etc: B0. Allow π	H_0: μ= 42, H_1: μ> 42: B1 only
			B(120, 0.35)	M1	B(120, 0.35) stated or implied	
			≈ N(42, 27.3)	M1	N(np, npq), their attempt at 120 × 0.35	120 × 0.35 × 0.65 *Not* N(np, nq).
			α: $z = \dfrac{49.5 - 42}{\sqrt{27.3}}$	A1ft	Standardise, with their np and $\sqrt{npq}$, right cc Allow both 49.5 and 50.5 and both in CR	$\sqrt{50}$ or $\sqrt{120}$: M1M1A0A0A1M0A0
			= 1.435	A1	z in range [1.43, 1.44] before rounding	Or p in range [0.075, 0.0764]
			> 1.282 [or 0.0757 < 0.1]	A1ft	Comparison with 1.282, ft on z/p or $\sqrt{120}$	Or p explicit comparison with 0.1
			β: CV = 42.5 + 1.282 × $\sqrt{27.3}$ [= 49.198]	A1ft	CV 42.5 + z × $\sqrt{27.3}$, ignore LH, ft on np, npq	No cc: 48.618, can get A0A1A0
			z = 1.282 and compare 50	A1	z = 1.282 used in RH CV and compare 50	
			CR ≥ 50 or ≥ 49.2	A1ft	CV correct ft on z, but don't worry about ≥	Must round up. 49 from 49.2: A1A1A0
			Reject H_0.	M1	Consistent first conclusion, needs correct method and comparison	Can give M1A1 even if comparison not explicit. Allow from exact binomial
			Significant evidence that proportion who know regulations has increased	A1ft	Contextualised, needs “who know regulations” or “pupils”, and “evidence”	Ft on TS & CV Or exact equivalent somewhere
			$np > 5$ [= 42] from normal attempted	M1	From p = 0.35 or 5/12, don't need 42	*or* n large or p close to 0.5 asserted
			$nq = 78 > 5$ and no others apart from n large	A1	Need 78, or 70 from 5/12, *not npq*	*and* the other qualitative reason asserted
			SC: If B0, B(120, 5/12): N(50, 29.17) M1M1 $np > 5$, $nq = 70 > 5$: M1A1 Max 4 SC: P(≥ 42): B2 M1M1A0A0A1M0A0		Wrong or no cc [1.627, 0.0519 or 1.5311, 0.0629]: loses (α) first two A1A1 only Exact B(120, 0.35): P(≥ 50) = 0.076824, CR ≥ 50. B2M1, M0A0A0A0, M1A1M0A0	

					NB: If S3 difference of proportions test used, consult PE	**Examiner's Comments** This hypothesis test involving a normal approximation to binomial was generally well done, apart from those who used the sample proportion 5/12 instead of the hypothesised proportion 0.35. The plan is to convert from B(120, 0.35) to N(42, 27.3) and either to find a critical value or to find P(≥ 50). Common causes of loss of marks were: • Omission of the continuity correction (42.5 for the CV, 49.5 for the probability) • Failure to justify the approximation fully (examiners needed to see $nq = 78$ if the condition $nq > 5$ was used, while "$npq > 5$" is wrong, as is "n large and $np > 5$") • Stating the hypotheses in terms of μ rather than the original parameter p • Attempts to use √120 or √50 in the standardisation.
			Total	11		
2		i	$\hat{\mu} = \bar{x} = 81$	B1	81 only, can be implied	
		i	$\dfrac{329800}{50} - 81^2 \qquad [= 35]$	M1	Correct formula for biased estimate, their "81", can be implied	
		i	$\times \dfrac{50}{49}; \qquad = 35.71$	M1	Multiply by 50/49. SC: single formula: M2, or M1 if wrong but divisor 49 anywhere [can be recovered if correctly done in part (ii)]	

		i		A1	A.r.t. 35.7 — can't be recovered from part (ii). Can be implied	
		i	$1-\Phi\left(\frac{90-81}{\sqrt{35.71}}\right)=1-\Phi(1.506)=1-0.9339$	M1	Standardise with their μ and σ, allow σ^2, cc but not √50	
		i	= 6.61% or 0.0661	A1	Answer, a.r.t. 6.6% or 0.066 **Examiner's Comments** It was perhaps indicative of candidates' over-rigid ways of answering questions that many omitted the $n/(n-1)$ factor for the variance in this part, yet went on and used it in the more familiar context of part (ii). More predictable was that many attempted to use a $\sqrt{n}$ factor in the standard deviation in this part, where it is wrong. However, the correct answer was often seen.	
		ii	H_0: $\mu = 80$ H_1: $\mu \neq 80$	B2	Correct, B2. One error, e.g. wrong or no symbol, >, B1, but x or $\bar{x}$ or t etc, or 81, B0. NB: If both hypotheses involve 81, *can't* get final M1	
		ii	$z=\frac{81-80}{\sqrt{35.71/50}}=1.183 \quad [or\ p=0.1183]$	M1	Standardise, with √50, allow √, sign or cc errors, allow from biased variance	
		ii		A1	z, a.r.t. 1.18, or p, a.r.t. 0.118. Allow −1.18.	
		ii	< 1.645	B1	Their $z < 1.645$ or $p > 0.05$, *not* if one–tail. Allow −1.18 > −1.645. *Not* just 1.645 seen.	
		ii	CV $80+1.645\sqrt{\frac{35.71}{50}}=81.39$	M1	80 + zs/√50, allow √ or cc errors, ignore – (no marks for – alone);	

		ii		B1	$z = 1.645$ used in this expression (not just seen), *not* from one–tail	
		ii	81 < 81.39	A1	Compare CV with 81, allow 81.08 from one–tailed ($z = 1.282$) (but not on their σ) SC: $81 - 1.645\sqrt{\frac{35.71}{50}}$: If H_0: μ = 80: (B2) M1B1A0M0A0. If H_0: μ = 81: (B0) M1B1A1 (79.61) M0A0	
		ii	Do not reject H_0.	M1	Correct first conclusion, needs √50, correct comparison type, μ and $\bar{x}$ not consistently wrong way round (thus H_0: μ = 81 can get B0 M1A1A1 M0A0, max 3/7) In method β, it needs to be clear that comparison involves $\bar{x}$.	
		ii	Insufficient evidence that the mean time is not 80 minutes.	A1ft	Contextualised (mention "time"), acknowledge uncertainty ("evidence that…") *Not* "significant evidence that mean time is 80" FT on wrong z-value or wrong critical value if previous mark gained SC: One–tailed: can get B1B0 M1A1B0 M1A1, max 5/7 No √50: can get B2 M0 B1 M0, max 3/7 **Examiner's Comments** Many answered this question very well, although relatively few achieved all 7 marks. Those who omitted the √50 factor here lost 4 marks, as did those who stated their hypotheses in terms of 81 and not 80 (a serious mistake emphasised in all recent Reports to Centres). Those who used a critical region often centred it on 81 rather than 80; were these S3 candidates who had confused the method with confidence intervals? The same	

					comments about the need to state the conclusion properly apply as in question 6(iii).	
		iii	(a) Yes (single observation only)	B1	No reason needed, but withhold if wrong reason seen. Allow "yes, no distn given" "No" *and* refer to central limit theorem or "large sample" {note for scoris zoning — (a) and (b) to be in single zone} **Examiner's Comments** As usual a question that tests understanding of the Central Limit Theorem was poorly answered. "No, yes" was more common than the correct "yes, no" (+ reason). Many said that you didn't have to assume a normal distribution in part (i) as *n* was large; clearly they had not realised that in part (i) we are talking about probabilities for a single observation. These candidates often gave "yes" as their answer to (b), presumably on no better grounds than expecting the two answers to be different. Another common wrong answer to (a) was "no as we know it is normal"; Examiners find it hard to account for the misconception here.	
		iii	(b) No, CLT applies to large sample	B1		
			Total	15		
3		i	$\bar{t} = 11.76$	B1	11.76 seen or implied	
		i	$\hat{\sigma}^2 = \frac{120}{119}\left(\frac{18737.712}{120} - 11.76^2\right) = 18$	M1	Biased estimate (= 17.85)	
		i		M1	× 120/119, *or* single formula with 119 divisor	i.e. correct single formula gets M2
		i		A1	Answer 18 ± 0.05	

		i	H_0: μ = 11.0, H_1: μ ≠ 11.0	B2	One error, B1, but $\bar{t}$, *t*, *x* etc: B0 (*u*: B1)	If both hypotheses involve 11.76, only
		i	α: $z = \dfrac{11.76 - 11.0}{\sqrt{18/120}} = \mathbf{1.9623}$	M1	Standardise with 120, ignore cc or √ errors	further mark possible is next M1 [max 5/11]
		i		A1	A.r.t. (±)1.96 *or p* ∈[0.0245, 0.025] www	120 omitted gets no further marks [max 6/11]
		i	> 1.645	A1	Compare explicitly with (±)1.645 or 0.05, consistent with their *z* or *p*. *[Needs to be "next to" TS]*	Ignore "N(11.76, …)" *unless* hypotheses omitted altogether, in which case treat as hypotheses in terms of 11.76
		i	β: CV 11.0 ± 1.645 × √ (18/120)	M1	11.0 + *z* σ/√120, needs 120 and + or ±	If 11.76 – *z* σ/√120, give M1A0A0 M0A0
		i	= 11.637 (or 10.363)	A1	Ignore 10.363	(even if correct hypotheses)
		i	11.76 > 11.64	A1	Explicit comparison, consistent tail	
		i	Reject H_0. Significant evidence that the average time has changed	M1	Correct first conclusion, allow "Accept H_1"	Needs correct method (including 120) and
		i		A1ft	Contextualised, acknowledge uncertainly, FT on wrong CR / *z* / *p*	comparison type, 11.0 in at least one hypothesis Allow "increase" instead of "change"
		i			**Examiner's Comments** Another standard question, if lengthy, and generally well answered. It is pleasing to note how few candidates gave their hypotheses in terms of the sample mean (H_0: *μ* = 11.76 instead of the correct H_0: *μ* = 11.0). Most, too, remembered to multiply the variance by 120/119. However, quite a few omitted the √120 in the denominator of the standardisation. Conclusions were well stated.	
		ii	No, the Central Limit Theorem applies	B1	or "No, large sample". Withhold if extra wrong or irrelevant reason(s) given **Examiner's Comments**	Needs both "no" and reason.

					As so often, a question that asked whether the normal distribution had to be assumed was met with a range of bafflingly self-contradictory answers. 'Yes because we can use the central limit theorem' was probably typical. Perhaps the misunderstanding stems from what the word assume means, perhaps from a failure to distinguish between the two different distributions in the question. The question asked whether the *consultation times* (that is, the parent population) had to be normal, whereas the calculation involves the *sample mean*. The distribution of the parent population does *not* have to be normal, because the central limit theorem tells us that the distribution of the sample mean *is* (approximately) normal.	
			Total	12		
4		i	H_0: μ = 0.55, H_1: μ < 13.3	B2	Both correct: B2. One error [e.g. *p*, ≠, no symbol] B1, but $\bar{x}$, $\bar{X}$ etc B0	
			α:			
		i	$z = \dfrac{12.48 - 13.3}{\sqrt{12.25/50}} = -1.6566\ [p = 0.0488]$	M1	Standardise with √50, allow √ errors, allow cc, allow 13.3 – 12.48	
		i		A1	*z* in range [–1.66, –1.65], or p in range [0.04875, 0.0489], allow 0.9512 only if consistent	
		i	[12.25/50 = 0.245] < –1.645 [*p* < 0.05]	B1	Compare with –1.645, allow +1.6566 with +1.645, or *p* with 0.05/0.95 as consistent	

			β:			
		i	CV $13.3-1.645\sqrt{\frac{12.25}{50}}=12.4857\ldots$	M1	13.3 – $z\sigma/\sqrt{50}$, any recognisable *z*, allow √errors etc, ignore 13.3 + …	
		i		B1	*z* = 1.645	
		i	12.48 < CV	A1	Compare 12.49 (or better) with 12.48, ignore 13.3 + … SC: 2-tailed, 12.33 gets B1B0 M1B0A1ft M1A1	
		i	Reject H_0.	M1	Consistent, needs √50, like-with-like comparison, hypotheses *not* 12.48 Contextualised, acknowledge uncertainty, their *z* SC1: 2-tailed: can get B1 M1A1B0 M1A1 max 5/7 SC2: No √50: can get B2 M0A0 B1 M0 max 3/7 SC3: $\bar{x}$ and μ confused consistently: can get B0 M1A1 B1 M0 SC4: 50/49 used in (i): can get B2 M1A0B1 M1A1 (6) in (i), M1 in (ii) **Examiner's Comments**	
		i	Significant evidence that animals in zoos have shorter expected lifetime	A1ft	This was by far the least well answered question, and between them the two parts produced often produced chaotic results. The correct method was that part (i) was a straightforward test for the mean of a normal distribution, using the given variance with a divisor of 50, while in part (ii) it was necessary to multiply the given variance by 50/49 (and then divide by 50 again). Unfortunately a lot of candidates did not appreciate the difference between the two variances and attempted somewhat desperately to find some other difference between parts (i) and (ii), usually dividing by 50 in one part but not the other. Some wrote	

					identical solutions to the two parts, which was at least honest! It may be worth spelling out that a divisor of *n* is always needed when the variable used for calculation is a sample mean. Several candidates took 12.25 to be the standard deviation, which of course led to very wrong numerical answers (though they could still get most of the marks). More pleasingly, only a small number of candidates began with the completely wrong hypotheses H_0: μ = 12.48, H_1: μ ≥ 12.48. Those who used the critical region method needed to be careful with accuracy as the critical value and sample mean differ only in the fourth significant figure; in fact it is always wise in this type of test to calculate critical values to plenty of decimal places. The fact that an apparently trifling change in the test produces the opposite conclusion is perhaps a commentary on the "significance level" approach to testing, and in the real world the use of *p*-values (which here would be 0.0488 and 0.0505) has become common. Although many gave their conclusions in an admirably correct way, a statement that "the mean lifetime of animals *has been reduced*" is wrong; candidates who wrote this were answering a different question.	
		ii	$\hat{\sigma}^2 = \frac{50}{49} \times 12.25 \quad [= 12.5]$	M1	Multiply 12.25 by 50/49, allow √ etc, allow if done in part (i) but then 0	
		ii	$z = \frac{12.48 - 13.3}{\sqrt{12.5/50}} = -1.64$ [p = 0.0505]	M1	Standardise with √50	

		ii		A1	Obtain a.r.t. −1.64, allow +1.64 if consistent with (i).	
		ii	> −1.645 [$p > 0.05$]	B1	Compare with same CV as in (i)	
		ii	Opposite conclusion	A1ft	State opposite conclusion (ft), any form, allow $\overline{x}$/μ here , needs M1M1 *Identical mark scheme for method β, CV 12.4775* SC1: 50 omitted consistently in both: M1M0A0B1A1 max 3/5 SC2: no √50 in (i), √50 but not 50/49 in (ii): M0M1A0B1A1 max 3/5 **Examiner's Comments** This was by far the least well answered question, and between them the two parts produced often produced chaotic results. The correct method was that part (i) was a straightforward test for the mean of a normal distribution, using the given variance with a divisor of 50, while in part (ii) it was necessary to multiply the given variance by 50/49 (and then divide by 50 again). Unfortunately a lot of candidates did not appreciate the difference between the two variances and attempted somewhat desperately to find some other difference between parts (i) and (ii), usually dividing by 50 in one part but not the other. Some wrote identical solutions to the two parts, which was at least honest! It may be worth spelling out that a divisor of *n* is always needed when the variable used for calculation is a sample mean. Several candidates took 12.25 to be the standard deviation, which of course led to very wrong numerical answers (though they could still get most of the marks).	

					More pleasingly, only a small number of candidates began with the completely wrong hypotheses H_0: $\mu = 12.48$, H_1: $\mu \geq 12.48$. Those who used the critical region method needed to be careful with accuracy as the critical value and sample mean differ only in the fourth significant figure; in fact it is always wise in this type of test to calculate critical values to plenty of decimal places. The fact that an apparently trifling change in the test produces the opposite conclusion is perhaps a commentary on the "significance level" approach to testing, and in the real world the use of *p*-values (which here would be 0.0488 and 0.0505) has become common. Although many gave their conclusions in an admirably correct way, a statement that "the mean lifetime of animals *has been reduced*" is wrong; candidates who wrote this were answering a different question.	
		iii	Yes as population not known to be normal	B1	Not "*n* large" unless "Yes, not known normal, but *n* large so can use" No wrong extras, e.g. "depends on whether it's sample or population" **Examiner's Comments** As so often, a question testing the use of the Central Limit Theorem revealed misunderstandings. As usual the question required *either* a necessary or a sufficient condition (here a necessary condition) and the mark scheme penalised the quotation of the wrong condition (though an answer such as "We need to use the CLT because the parent distribution is not stated	

					to be normal, but we can use it as n is large" was accepted, the key word being "can"). There seems to be at least two widespread misunderstandings about the CLT. One is that a large sample makes the parent distribution normally distributed; put like this it is obviously wrong. Another is expressed by the answer "We do not need to use the Central Limit Theorem as it is a large sample" (or "a continuous distribution"); what do these candidates think that the CLT actually says? It may be worth emphasising that we are talking about two different variables (a single observation X, and the mean of n observations $\bar{X}$), and that these two variables have different distributions. The statement of the CLT is that it does not matter what the distribution of a single observation is; if the sample size is large enough, the distribution of the sample mean $\bar{X}$ s approximately normal.	
			Total	13		
5			$H_0 : \mu = 32.5$ $H_1 : \mu \neq 32.5$ where μ is mean time spent by all customers $X \sim N\left(32.5, \frac{8.2^2}{50}\right)$ and $X > 34.5$ $P(X > 34.5) = 0.0423$	B1(AO1.1) B1(AO2.5) M1(AO3.3) A1(AO3.4) A1(AO1.1) M1(AO1.1)	Must be stated in terms of parameter Values **B1B0** for one error, e.g. undefined μ or 1-tail Stated or implied | Use of 34.5 **B0B0** **OR** **M1**	

			Comparison with 0.025 Do not reject H_0 Insufficient evidence that mean time in the library has changed	A1FT(AO2.2b) [7]	BC Allow comparison with 0.05 if H1 : $\mu > 32.5$ In context, not definite; FT their 0.0423, but not comparison with 0.05	$\frac{34.5-32.5}{8.2\div\sqrt{50}}$ allow without square root **A1** = 1.725 **A1** Comparison with 1.96 (allow comparison with 1.645 if H1 : $\mu > 32.5$) FT their 1.725, but not comparison with 1.645
			Total	7		
6		a	$X \sim \mathrm{N}(15.5, \frac{2.6^2}{12})$ $15.5 + 1.645\frac{2.6}{\sqrt{12}}$ Critical region is $\bar{x} > 16.7$ (3 sf)	M1(AO 3.3) M1(AO 3.4) A1(AO 1.1) [3]	stated or implied	

		b	H_0: $\mu = 15.5$ H_1: $\mu > 15.5$ where μ is mean time by all employees $\bar{x} = 16.8$ 16.8 is within CR		ft their $\bar{x}$ & CR Reject H_0. There is evidence that mean time for task is greater than 15.5 (minutes)	B1(AO1.1) B1(AO2.5) A1ft(AO3.3) M1(AO1.1) A1ft(AO2.2b) [5]	In terms of parameter values **B1B0** one error eg undefined μ or two-tail Use of 17.5 B0B0 **OR** $P(X > 16.8) = 0.0416$ (3 sf) Comp 0.05 Allow 0.25 if H_1: $\mu \neq 15.5$ In context, not definite. ft their 0.0416 but not comp with 0.25	
			Total	8		

7		a	H_0: $\mu = 10.5$ where μ is pop mean time in shop H_1: $\mu \neq 10.5$ $\bar{X} \sim N(10.5, \frac{4.2}{\sqrt{50}})$ and $X = 12$ $P(\bar{X} > 12) = 0.00578$ or better Compare with 0.005 Do not reject H0 Insufficient evidence that mean time has changed	B1 (AO1.1) B1 (AO2.5) M1 (AO3.3) A1 (AO3.4) M1 (AO1.1) M1 (AO1.1) A1 (AO2.2b) [7]	One error, eg undefined μ or 1-tail: **B0B1** May be implied or 0.006 or 0.0058 **BC** In context. Not definite, eg "Mean time has not changed" **A0**	
		b	0.01	B1 (AO1.2) [1]		
			Total	8		