1. A large company surveyed its staff to investigate the awareness of company policy. The company employs 6000 full time staff and 4000 part time staff.

(a) Describe how a stratified sample of 200 staff could be taken.

(3)

(b) Explain an advantage of using a stratified sample rather than a simple random sample.

(1)

A random sample of 80 full time staff and an independent random sample of 80 part time staff were given a test of policy awareness. The results are summarised in the table below.

	Mean score ($\bar{x}$)	Variance of scores (s^2)
Full time staff	52	21
Part time staff	50	19

(c) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean policy awareness scores for full time and part time staff are different.

(7)

(d) Explain the significance of the Central Limit Theorem to the test in part (c).

(2)

(e) State an assumption you have made in carrying out the test in part (c).

(1)

After all the staff had completed a training course the 80 full time staff and the 80 part time staff were given another test of policy awareness. The value of the test statistic z was 2.53

(f) Comment on the awareness of company policy for the full time and part time staff in light of this result. Use a 1% level of significance.

(2)

(g) Interpret your answers to part (c) and part (f).

(1)
(Total 17 marks)

2. The lengths of a random sample of 120 limpets taken from the upper shore of a beach had a mean of 4.97 cm and a standard deviation of 0.42 cm. The lengths of a second random sample of 150 limpets taken from the lower shore of the same beach had a mean of 5.05 cm and a standard deviation of 0.67 cm.

(a) Test, using a 5% level of significance, whether or not the mean length of limpets from the upper shore is less than the mean length of limpets from the lower shore. State your hypotheses clearly.

(8)

(b) State two assumptions you made in carrying out the test in part (a).

(2)
(Total 10 marks)

3. A sociologist is studying how much junk food teenagers eat. A random sample of 100 female teenagers and an independent random sample of 200 male teenagers were asked to estimate what their weekly expenditure on junk food was. The results are summarized below.

	n	mean	s.d
Female teenagers	100	£5.48	£3.62
Male teenagers	200	£6.86	£4.51

(a) Using a 5% significance level, test whether or not there is a difference in the mean amounts spent on junk food by male teenagers and female teenagers. State your hypotheses clearly.

(7)

(b) Explain briefly the importance of the central limit theorem in this problem.

(1)
(Total 8 marks)

4. The time, in minutes, it takes Robert to complete the puzzle in his morning newspaper each day is normally distributed with mean 18 and standard deviation 3. After taking a holiday, Robert records the times taken to complete a random sample of 15 puzzles and he finds that the mean time is 16.5 minutes. You may assume that the holiday has not changed the standard deviation of times taken to complete the puzzle.

Stating your hypotheses clearly test, at the 5% level of significance, whether or not there has been a reduction in the mean time Robert takes to complete the puzzle.

(Total 7 marks)

5. In a trial of diet *A* a random sample of 80 participants were asked to record their weight loss, x kg, after their first week of using the diet. The results are summarised by

$$\sum x = 361.6 \quad \text{and} \quad \sum x^2 = 1753.95$$

(a) Find unbiased estimates for the mean and variance of weight lost after the first week of using diet *A*.

(5)

The designers of diet *A* believe it can achieve a greater mean weight loss after the first week than a standard diet *B*. A random sample of 60 people used diet *B*. After the first week they had achieved a mean weight loss of 4.06 kg, with an unbiased estimate of variance of weight loss of 2.50 kg^2.

(b) Test, at the 5% level of significance, whether or not the mean weight loss after the first week using diet *A* is greater than that using diet *B*. State your hypotheses clearly.

(7)

(c) Explain the significance of the central limit theorem to the test in part (b).

(1)

(d) State an assumption you have made in carrying out the test in part (b).

(1)

(Total 14 marks)

6. A biologist investigated whether or not the diet of chickens influenced the amount of cholesterol in their eggs. The cholesterol content of 70 eggs selected at random from chickens fed diet A had a mean value of 198 mg and a standard deviation of 47 mg. A random sample of 90 eggs from chickens fed diet B had a mean cholesterol content of 201 mg and a standard deviation of 23 mg.

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not there is a difference between the mean cholesterol content of eggs laid by chickens fed on these two diets.

(7)

(b) State, in the context of this question, an assumption you have made in carrying out the test in part (a).

(2)
(Total 9 marks)

7. Upon entering a school, a random sample of eighty girls and an independent random sample of eighty boys were given the same examination in mathematics.
The girls and boys were then taught in separate classes.
After one year, they were all given another common examination in mathematics.

The means and standard deviations of the boys' and the girls' marks are shown in the table.

Examination marks				
	Upon entry		After 1 year	
	Mean	Standard deviation	Mean	Standard deviation
Boys	50	12	59	6
Girls	53	12	62	6

You may assume that the test results are normally distributed.

(a) Test, at the 5% level of significance, whether or not the difference between the means of the boys' and girls' results was significant when they entered the school.

(7)

(b) Test, at the 5% level of significance, whether or not the mean mark of the boys is significantly less than the mean mark of the girls in the 'After 1 year' examination.

(5)

(c) Interpret the results found in part (a) and part (b).

(1)
(Total 13 marks)

8. It is known from past evidence that the weight of coffee dispensed into jars by machine *A* is normally distributed with mean μ_A and standard deviation 2.5 g. Machine *B* is known to dispense the same nominal weight of coffee into jars with mean μ_B and standard deviation 2.3 g. A random sample of 10 jars filled by machine *A* contained a mean weight of 249 g of coffee. A random sample of 15 jars filled by machine *B* contained a mean weight of 251 g.

(a) Test, at the 5% level of significance, whether or not there is evidence that the population mean weight dispensed by machine *B* is greater than that of machine *A*.

(7)

(b) Write down an assumption needed to carry out this test.

(1)
(Total 8 marks)

9. A scientist monitored the levels of river pollution near a factory. Before the factory was closed down she took 100 random samples of water from different parts of the river and found an average weight of pollutants of 10 mg l^{-1} with a standard deviation of 2.64 mg l^{-1}. After the factory was closed down the scientist collected a further 120 random samples and found that they contained 8 mg l^{-1} of pollutants on average with a standard deviation of 1.94 mg l^{-1}.

Test, at the 5% level of significance, whether or not the mean river pollution fell after the factory closed down.

(Total 11 marks)

1. (a) Label full time staff 1 – 6000, part time staff 1 – 4000 M1

Use random numbers to select M1

Simple random sample of 120 full time staff and 80 part time staff A1 3

<u>Note</u>

1st M1 for attempt at labelling full-time and part-time staff. One set of correct numbers.

2nd M1 for mentioning use of random numbers

1st A1 for s.r.s. of 120 full-time and 80 part-time

(b) Enables estimation of statistics / errors for each strata <u>or</u> "reduce variability" <u>or</u> "more representative" <u>or</u> "reflects population structure" **NOT** "more accurate" B1 1

(c) H_0: $\mu_f = \mu_p$, H_1: $\mu_f \neq \mu_p$ (accept μ_1, μ_2) B1

s.e. = $\sqrt{\frac{21}{80}+\frac{19}{80}}$, $z = \frac{52-50}{\sqrt{\frac{21}{80}+\frac{19}{80}}} = \left(2\sqrt{2}\right)$ M1, M1

= 2.828… (awrt **2.83**) A1

Two tailed critical value z = 2.5758 (or prob of awrt 0.002 (<0.005) or 0.004 (<0.01)) [2.828 > 2.5758 so] significant evidence to reject H_0 B1 dM1

There is evidence of a difference in policy awareness between full time and part time staff A1ft 7

<u>Note</u>

1st M1 for attempt at s.e. – condone one number wrong. NB correct s.e. = $\sqrt{\frac{1}{2}}$

2nd M1 for using their s.e. in correct formula for test statistic. Must be $\frac{\pm(52-50)}{\sqrt{\frac{p}{q}+\frac{r}{s}}}$

3rd dM1 **dep. on 2nd M1** for a correct statement based on their normal cv and their test statistic

2nd A1 for correct comment in context. Must mention "scores" or "policy awareness" and types of "staff". Award **A0** for a one-tailed comment. Allow ft

(d) Can use mean full time and mean part time B1
~ Normal B1 2

<u>Note</u>

1st B1 for mention of mean(s) <u>or</u> use of $\bar{X}$, provided $\bar{X}$ clearly refers to full-time or part-time

2nd B1 for stating that distribution can be assumed normal
e.g. "mean score of the test is normally distributed" gets B1B1

(e) Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population B1 1

(f) 2.53 < 2.5758, not significant or do not reject H_0 M1
So there is insufficient evidence of a difference in mean awareness A1ft 2

Note

M1 for correct statement (may be implied by correct contextualised comment)
A1 for correct contextualised comment. Accept "no difference in mean scores". Allow ft

(g) Training course has closed the gap between full time staff and part time staff's mean awareness of company policy. B1 1

Note

B1 for correct comment in context that implies training was effective. This must be supported by their (c) and (f). Condone one-tailed comment here.

[17]

2. (a) μ_U ~ mean length of upper shore limpets, μ_L ~ mean length of lower shore limpets

$H_0 : \mu_U = \mu_L$ both B1

$H_0 : \mu_U < \mu_L$

s.e. = $\sqrt{\dfrac{0.42^2}{120} + \dfrac{0.67^2}{150}}$ M1

= 0.0668

$z = \dfrac{5.05 - 4.97}{0.0668} = (\pm)1.1975$ awrt $\pm$ **<u>1.20</u>** dM1 A1

Critical region is $z \geq 1.6449$, or probability = awrt (0.115 or 0.116) z = $\pm$ 1.6449 B1

(1.1975 < 1.6449) therefore not in critical region / accept H_0/not significant (or $P(Z \geq 1.1975)$ =0.1151, 0.1151>0.05 or z not in critical region) M1

There is no evidence that the limpets on the upper shore are shorter than the limpets on the lower shore. A1 8

Assume the populations or variables are independent

Note

1st B1	If μ_1, μ_2 used then it must be clear which refers to upper shore. Accept sensible choice of letters such as *u* and *l*.
1st M1	Condone minor slips e.g. $\frac{0.67^2}{120}$ or $\frac{0.67}{120}+\frac{0.42^2}{120}$ etc i.e. swapped *n* or one sd and one variance but M0 for $\sqrt{\frac{0.67}{150}+\frac{0.42}{120}}$
1st A1	can be scored for a fully correct expression. May be implied by awrt 1.20
2nd dM1	is dependent upon the 1st M1 but can ft their se value if this mark is scored.
2nd A1	for awrt ($\pm$) 1.20
3rd M1	for a correct statement based on their *z* value and their cv. No cv is M0A0 If using probability they must compare their *p*(<0.5) with 0.05 (o.e) so can allow 0.884< 0.95 to score this 3rd M1 mark. May be implied by their contextual statement and M1A0 is possible.

(b) Standard deviation of sample = standard deviation of population B1

[Mention of Central Limit Theorem does NOT score the mark] B1 2

Note

3rd A1	for a correct comment to accept null hypothesis that mentions length of limpets on the two shores.
1st B1	for one correct statement. Accept "samples are independent"
2nd B1	for both statements

[10]

3. (a) $H_0: \mu_f = \mu_F$ $H_1: \mu_F \neq \mu_M$ (Allow μ_1 and μ_2) B1

$z = \dfrac{6.86 - 5.48}{\sqrt{\dfrac{4.51^2}{200} + \dfrac{3.62^2}{100}}}$ M1A1

= 2.860… awrt (±)**2.86** A1

2 tail 5% critical value (±) 1.96(or probability awrt 0.0021~0.0022) B1

Significant result or reject the null hypothesis (o.e.) M1

There is evidence of a difference in the (mean) amount spent on junk food by male and female teenagers A1ft 7

1st M1 for an attempt at $\dfrac{a-b}{\sqrt{\dfrac{c}{100 \text{ or } 200} + \dfrac{d}{100 \text{ or } 200}}}$ with 3 of *a*, *b*, *c* or *d* correct

1st A1 for a fully correct expression

2nd B1 for ± 1.96 but only if their H_1 is two-tail (it may be in words so B0B1 is OK)
If H_1 is one-tail this is automatically B0 too.

2nd M1 for a correct statement based on comparison of their *z* with their cv.
May be implied

3rd A1 for a correct conclusion in context based on their *z* and 1.96.
Must mention junk food or money and male vs female.

(b) CLT enables us to assume $\overline{F}$ and $\overline{M}$ are normally distributed B1 1

B1 for $\overline{F}$ or $\overline{M}$ mentioned. Allow "mean (amount spent on junk food) is normally distributed"
Read the whole statement e.g. " original distribution is normal so mean is…" scores B0

[8]

4. $H_0: \mu = 18$, $H_1: \mu < 18$ B1,B1

$z = \dfrac{16.5 - 18}{3/\sqrt{15}}$ =,–1.9364... AWRT –1.94 M1,A1

5% one tail c.v. is z = (–) 1.6449 or probability (AWRT 0.026) (±) 1.6449 B1

–1.94 < –1.6449 or significant or reject H_0 or in critical region M1

There is evidence that the (mean) time to complete the puzzles has reduced

Or Robert is getting faster (at doing the puzzles) A1ft 7

1st & 2nd B1 must see μ and 18

1st M1 for attempting test statistic, allow ±. Or attempt at critical value for $\overline{X} : \mu - z \times \frac{3}{\sqrt{15}}$

1st A1 for AWRT – 1.94. Allow use of $|z|$ = +1.94 to score M1A1.
Or critical values = AWRT 16.7.

3rd B1 for AWRT 0.026 (i.e. correct probability only) or ±1.6449.
(May be seen in cv formula)

2nd M1 for correct comparison or statement relating their test statistic and 1.6449 or their probability and 0.05. Ignore their hypotheses if any or assume they were correct.

2nd A1ft for conclusion in context which refers to "speed" or "time". Depends only on previous M

[7]

5. (a) $\hat{\mu} = \overline{x} = \frac{361.6}{80}$, = <u>4.52</u> M1, A1

$\hat{\sigma}^2 = s^2 = \frac{1753.95 - 80 \times \overline{x}^2}{79}$ = (1.51288...) M1A1ft

AWRT <u>1.51</u> A1 5

2nd M1 for a correct attempt at s or s^2, A1ft for correct expression for s^2, ft their mean.

N.B. $\sigma_n^2 = 1.49...$ so $\frac{80}{79} \times 1.49...$ is M1A1ft

(b) H_0: $\mu_A = \mu_B$ H_1: $\mu_A > \mu_B$ B1B1

Denominator M1

$$z = \frac{4.52 - 4.06}{\sqrt{\frac{1.51...}{80} + \frac{2.50}{60}}} = \left(\frac{0.46}{\sqrt{0.0605...}}\right)$$ z dM1

= (±) 1.8689... AWRT (±) <u>1.87</u> A1

One tail c.v. is z = 1.6449
(AWRT 1.645 or probability AWRT 0.0307 or 0.0308) B1

(significant) there is evidence that diet A is better than diet B <u>or</u> evidence that (mean) weight lost in first week using diet A is more than with B A1ft 7

1st B1 can be given for $\mu_1 = \mu_2$, but 2nd B1 must specify which is *A* or *B*.

1st M1 for the denominator, follow through their 1.51. Must have square root can condone 2.50^2 but $\sqrt{\frac{1.51^2}{80}+\frac{2.50^2}{60}}$ is M0.
Allow $\sqrt{\frac{1.51}{79}+\frac{2.50}{59}}$ leading to AWRT 1.85 to score M1M1A0 in (b) and can score in (d).

2nd dM1 for attempting the correct test statistic, dependent on denominator mark

1st A1 for AWRT ±1.87, may be implied by a correct probability.

2nd A1ft ft their test statistic vs their cv **only if** H_1 is correct and both Ms are scored

(c) CLT enables you to assume that $\bar{A}$ and $\bar{B}$ are normally distributed B1 1

B1 for stating <u>either</u> $\bar{A}$ or $\bar{B}$ (but not *A* or *B*) are normally distributed

(d) Assumed $\sigma_A^2 = s_A^2$ and $\sigma_B^2 = s_B^2$ (either) B1 1

B1 for either, can be stated in words in terms of variances or standard deviations.

[14]

6. (a) $H_0 : \mu_A = \mu_B,\ H_1 : \mu_A \neq \mu_B$ μ_A, μ_B OK both B1

$\text{SE} = \sqrt{\frac{47^2}{70}+\frac{23^2}{90}} (= \sqrt{37.43492...})$ M1 A1

Test statistic is $\pm\frac{198-201}{\text{SE}} = \pm 0.4903$ awrt ± 0.49 M1 A1

M1A1 probab awrt ± 0.312
B1 probab cv 0.025

$\text{cv} = (\pm)1.96$ B1

Insufficient evidence to reject H_0,
No significant difference between the mean cholesterol content of the two samples. A1ft 7

(require correct comparison for ft) context required.

(b)
- require 1 egg from each of 70 chickens of diet A to ensure independence, similarly for diet B
- no chickens in common between the two samples to ensure independence
- not same chickens on diet A and diet B because if it were we need to do a paired analysis — B1 B1 — 2

any one

[9]

7. (a) μ_b = mean mark of boys, μ_g = mean mark of girls.

$H_0 : \mu_b = \mu_g$ — both B1
$H_1 : \mu_b \neq \mu_g$

$$z = \frac{53 - 50}{\sqrt{\frac{144}{80} + \frac{144}{80}}}$$ M1 A1

$= 1.58$ — A1

Critical region $z \geq 1.96$ — B1
$1.58 < 1.96$ insufficient evidence to reject H_0. — M1
No diff. between mean scores of boys and girls. — A1 — 7

(b) $H_0 : \mu_b = \mu_g$
$H_1 : \mu_b < \mu_g$ — B1

$$z = \frac{62 - 59}{\sqrt{\frac{36}{80} + \frac{36}{80}}}$$ M1

$= 3.16$ — A1

Critical region $z \geq 1.6449$ (accept 1.645) — B1
$3.16 > 1.6449$ sufficient evidence to reject H_0. — A1 — 5
the mean mark for boys is less than the mean mark of the girls.

(c) Girls have improved more than boys
or girls performed better than boys after 1 year — B1 — 1

[13]

8. (a) $H_0 : \mu_A = \mu_B$; $H_1 : \mu_B > \mu_A$ B1

both and μ

$$z = \pm\frac{249 - 251}{\sqrt{\frac{2.5^2}{10} + \frac{2.3^2}{15}}}$$ M1 A1

249.251 accept $\sqrt{\frac{2.5}{10} + \frac{2.3}{15}}$ *for M*

$= \pm 2.0227\ldots$ A1

awrt ± 2.02

CV = ±1.6449
or P(Z ≥ 2.02) = 0.0212 – 0.0217, B1
or P(Z ⩽ 2.02) = 0.9788 – 0.9783

– 2.0227 < – 1.6449 or 2.0227 > 1.6449,
or 0.0212 – 0.0217 < 0.05 M1
or 0.9788 – 0.9783 > 0.95

comparison and consistency needed

There is evidence that the <u>mean amount of coffee</u> dispensed by B <u>is greater</u> than A. A1ft 7

context

(b) Machine B amounts are normally distributed. B1 1

[8]

9. μ_a and μ_b are mean weight of population a fter and before closure respectively. B1

H_0: $\mu_b = \mu_a$ B1 B1

H_1: $\mu_b > \mu_a$

$z = \dfrac{10-8}{\sqrt{\dfrac{2.64^2}{100}+\dfrac{1.94^2}{120}}}$ Fraction, denom Ok alone $\dfrac{\text{M1A1}}{\text{M1A1}}$

$z = \dfrac{2}{\sqrt{0.1011}} = 6.29$ awrt 6.29 A1

Critical region is $z \geq 1.6449$, $6.29 > 1.6449$ or in critical region or Reject H_0

(or $P(Z \geq 6.29) = 0$, $0 < 0.05$ or z in critical region or Reject H_0 B1M1)

1.6449 B1, M1

There is evidence that closing the factory has reduced mean river pollution A1∫ 11

[11]

1. Most candidates knew how to take a stratified sample by taking simple random samples in each stratum but they often forgot to describe how to label the members of the strata.

In (b) the commonest correct response was about the sample being more representative of the population but some missed the point and simply said that stratified sampling was “easier”.

The calculation in part (c) was carried out very well by most candidates. There were few errors with the standard error and most correctly concluded that there was evidence of a difference in policy awareness between the types of staff..

In part (d) most knew that the Central Limit Theorem had something to do with the normal distribution but they did not mention that it was the mean scores of full time and part time staff that can be assumed to be normally distributed.

There were some correct responses to part (e) but many just mentioned independence despite this being given in the stem to part (c) of the question.

Most gave a correct conclusion in part (f) and some correctly inferred in the final part that the training course had been effective.

Some had the correct idea in part (g) although their conclusions went further than the evidence suggested: they claimed that the scores of the part time staff had increased, which may well be the case, but the evidence presented was only sufficient to conclude that the “gap” between policy awareness of the types of staff has been closed.

2. Part (a) was usually answered very well with most stating their conclusion in context but a few losing a mark for simply using μ_1 and μ_2 in their hypotheses without giving any indication which population was which. In part (b) the Central Limit Theorem was a popular wrong answer (it is not an assumption but a theorem that is invoked because of the large samples) but many did mention independence and some that the sample variances were assumed equal to the population variances but it was rare to see a candidate earn both marks here.

3. Part (a) was answered well with only a few candidate confusing standard deviations with variances in the test statistic or using a critical value of 1.6449 instead of 1.96. Most had the hypotheses correct too with some giving them in terms of words and the population parameters. The conclusion was usually correct and in context. Part (b) exposed some candidate’s weak understanding of the central limit theorem with several lengthy answers failing to mention “mean” at all.

4. This should have been a straightforward question on a single-sample test for the mean but a large number of candidates scored poorly here. The first problem was the hypotheses; many stated these in terms of the mean before the holiday and the mean after the holiday as though a two–sample test was required. Those who did use the value 18 usually realised that a one-tailed test was required, the word “reduction” in the last line was the clue, and □ was usually used. The test statistic should have given a value of – 1.94 and some candidates lost a mark for losing the minus sign although of course a modulus test was acceptable. Those candidates who thought a two–sample test was required often had a curious test statistic based on the difference between the two normal distributions $R \sim N(18, 9)$ and $\bar{R} \sim N(16.5, \frac{9}{15})$. Despite these difficulties most candidates were able to give a correct conclusion to their test.

5. Part (a) was answered very well with only a small minority of candidates using a biased estimator of variance. The test was carried out quite well too, certainly with more success than that in Q3, but a number of candidates lost marks for confusing standard deviation with variance in the denominator of their test statistic. Parts (c) and (d) were not answered well. In part (c) surprisingly few candidates emphasised that the central limit theorem enabled one to assume that $\bar{A}$ and $\bar{B}$, rather than A and B, were normally distributed, where A and B represent the weight loss using diet A and diet B respectively. Very few noted the assumption that $\sigma_A^2 = s_A^2$ and $\sigma_B^2 = s_B^2$ in part (d).

6. Part (a) was well answered, but hypotheses sometimes appeared in ill thought out words rather than symbols. Part (b) was poorly answered with hardly any candidates going beyond independence and most ignoring the need for context beyond this. Given that two marks were on offer it was surprising that most candidates did not realise the need for a little more thought and a more detailed response.

7. The basic technique for testing for a difference between two means was well known and this question was usually answered well. The calculations were usually carried out correctly and the appropriate critical values used. Conclusions were usually given in context but in part (c) candidates often simply restated their conclusion from part (b) rather than attempting to summarise the overall findings by stating that the girls had improved more than the boys. In part (b) some candidates simply used μ_1 and μ_2 in their hypotheses and it was not possible to tell which referred to boys and which to the girls.

8. This question was quite well done but often marks were lost through lack of attention to detail. Rarely was any comparison of the z values or the probabilities shown and those who did were unable to involve sufficient context in their conclusions. Part (b) was badly done or omitted. Many candidates showed a lack of understanding in trying to use the Central Limit Theorem to give the final assumptions, despite the small sample sizes.

9. Again, candidates produced many excellent solutions. Some weaker candidates applied the square root in the denominator incorrectly, but full marks were not unusual here.