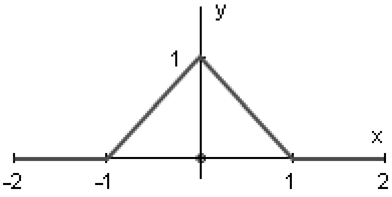
Q1	$E \sim \mathrm{N}(406,\ 12^2)$ When a candidate's answers suggest that (s)he appears to have neglected to use the difference columns of the Normal distribution tables penalise the first occurrence only.			
(i)	$\mathrm{P}(E < 420) = \mathrm{P}\left(Z < \frac{420-406}{12} = 1.1666\right)$ $= 0{\cdot}8783/4$	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	3
(ii)	$C \sim \mathrm{N}(406 \times 14.6 = 5927.6,$ $\sigma^2 = 12^2 \times 14.6^2 = 30695.04)$ P(this > 6000) = $\mathrm{P}\left(Z > \frac{6000-5927.6}{175.2} = 0.4132\right) = 1 - 0\cdot6602 = 0.3398$	 B1 B1 A1	Accept equivalent in £. Mean. Variance. Accept sd (= 175.2). Accept P(E > 6000/14.6) o.e. c.a.o.	3
(iii)	$B = C_1 + C_2 + C_3 \sim \mathrm{N}(17782.8,$ $\sigma^2 = 175.2^2 + 175.2^2 + 175.2^2 = 92085.12)$ Require b s.t. P(B < 100b) = 0.99 $\therefore \frac{100b - 17782.8}{303.455} = 2.326$ $\therefore 100b = 17782.8 + 2.326 \times 303.455 = 18488.6...$ (p) b = £184.89	 B1 B1 B1 A1	Accept equivalent in £, or E_1+ E_2+ E_3. Mean. ft from (ii). Variance. Accept sd (= 303.455…). ft from (ii). Accept P(E_1+ E_2+ E_3 < 100b/14.6) o.e. 2.326 seen. c.a.o. (Minimum 4 s.f. required in final answer.)	4
(iv)	$\mathrm{H_0}$: μ = 432 $\mathrm{H_1}$: μ < 432 where μ is the mean amount of electricity used. $\bar{x} = 422.16...$ $s_{n-1} = 13.075(4)$ Test statistic is $\frac{422.16-432}{\frac{13.075}{\sqrt{6}}}$ $= -1.842(13)$. Refer to t_5. Single-tailed 5% point is –2.015. Not significant. Insufficient evidence to suggest that the amount of electricity used has decreased on average.	B1 B1 B1 M1 A1 M1 A1 A1 A1	Both hypotheses. Hypotheses in words only must include "population". For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\bar{X} = ...$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean. s_n = 11.936 but do NOT allow this here or in construction of test statistic, but FT from there. Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 432 + (c's –2.015) × $13.075/\sqrt{6}$ (= 421.24) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –2.015) × $13.075/\sqrt{6}$ (= 432.92) for comparison with 432.) c.a.o. but ft from here in any case if wrong. Use of $\mu - \bar{x}$ scores M1A0. No ft from here if wrong. P(t < –1.842(13)) = 0.0624. Must be minus 2.015 unless absolute values are being compared. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Conclusion in context to include "on average" o.e.	9
				19

Q2				
(a) (i)	There are identifiable subgroups or strata that might exhibit different characteristics. Each stratum is randomly sampled. Use it to obtain a representative sample. Can get information on the individual strata.	E1 E1 E1 E1		4
(ii)	For each stratum $\ldots \times \frac{2000}{79368}$ giving 813.9, 836.9, 245.4, 103.8 so 814, 837, 245, 104	M1 A1	 All correct.	2
(b) (i)	The <u>population</u> (or underlying distribution) is assumed to be <u>symmetrical</u> about its <u>median</u>.	E2	E2, 1, 0. Award E1 for 2 out of 3 of the key features.	2
(ii)	H_0: $m = 0$ H_1: $m \neq 0$ where m is the population median difference for the percentages.	B1 B1	Both hypotheses. Hypotheses in words only must include "population". For adequate verbal definition.	

Diff	−0.66	0.02	−0.80	−0.91	0.28	0.76	0.40	1.68	−0.07	1.12
Rank	5	1	7	8	3	6	4	10	2	9

		M1	For differences. ZERO (out of 8) in this section if paired differences not used.	
		M1	For ranks.	
		A1	ft from here if ranks wrong.	
	$W_- = 2 + 5 + 7 + 8 = 22$	B1	(or $W_+ = 1 + 3 + 4 + 6 + 9 + 10 = 33$)	
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 10$.	M1	No ft from here if wrong.	
	Lower (or upper if 33 used) 5% tail is 10 (or 45 if 33 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c's test statistic.	
	No evidence to suggest a change in spending on average.	A1	ft only c's test statistic. Conclusion in context to include "on average" o.e.	10
				18

Q3				
(i)	Using mid- intervals 1.5, 1.7, etc $\bar{x} = \frac{205}{100} = 2.05$ $s = \sqrt{\frac{425.16 - 100 \times 2.05^2}{99}} = 0.2227(01...)$	M1 A1 E1	 Mean. s.d. Answer given; must show convincingly.	 3
(ii)	$f = 100 \times \mathrm{P}(1.8 \le M < 2.0)$ $= 100 \times \mathrm{P}(-1.1226 \le z < -0.2245)$ $= 100 \times ((1 - 0.5888) - (1 - 0.8691))$ $= 100 \times (0.4112 - 0.1309) = 28.03$	M1 A1 A1	Probability × 100. Correct Normal probabilities. ft c's mean. Must show convincingly using Normal distribution. ft c's mean.	 3
(iii)	H_0: The Normal model fits the data. H_1: The Normal model does not fit the data. $X^2 = 0.7294 + 0.1384 + 1.9623 + 3.5155 + 0.2437$ $= 6.589(3)$ Refer to χ^2_2. Upper 5% point is 5.991. Significant. Evidence suggests that the model does not fit the data.	B1 B1 M1 M1 A1 M1 A1 A1 A1	Ignore any reference to parameters. Merge first 2 and last 2 cells. Calculation of X^2. c.a.o. Allow correct df (= cells – 3) from wrongly grouped table and ft. Otherwise, no ft if wrong. $\mathrm{P}(X^2 > 6.589) = 0.0371$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Conclusion in context.	 9
(iv)	The model • overestimates in the 2.2 – 2.4 class, • underestimates in the 2 – 2.2 class. At lower significance levels the test would not have been significant.	 E1 E1 E1		 3
				18

Q4				
(i)		G1 G1 G1	One (straight) line segment correct. Second (straight) line segment correct. Fully labelled intercepts + no spurious other lines.	3
(ii)	E(X) = 0 (By symmetry.) $\text{E}(X^2) = \int_{-1}^{0} x^2(1+x)\text{d}x + \int_{0}^{1} x^2(1-x)\text{d}x$ $= \left[\frac{x^3}{3} + \frac{x^4}{4}\right]_{-1}^{0} + \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_{0}^{1}$ $= 0 - \left(\frac{-1}{3} + \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) - 0$ $= \frac{1}{6}$ $\therefore \text{Var}(X) = \frac{1}{6}\left(-0^2\right) = \frac{1}{6}$	B1 M1 M1 M1 A1	 One correct integral with limits (which may be implied subsequently). Second integral correct (with limits) or allow use of symmetry. Correctly integrated and attempt to use limits. c.a.o. Condone absence of explicit evidence of use of Var(X) = E(X^2) – E(X)^2.	5
(iii)	$\bar{L} \sim N\left(k, \frac{1}{300}\right)$ Normal distribution because of the Central Limit Theorem.	B1 B1 B1 E1	Normal. Mean. Variance. ft c's variance in (ii) (> 0) / 50. Any reference to the CLT.	4
(iv)	CI is given by $90.06 \pm$ 1.96 $\times \frac{1}{\sqrt{300}}$ $= 90.06 \pm 0.11316 = (89.947, 90.173)$	M1 B1 M1 A1	 ft c's variance in (ii) (> 0) / 50. Must be expressed as an interval.	4
(v)	It is reasonable, because 90 lies within the interval found in (iv).	E1	Or equivalent.	1
				17