Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Subsidiary

Wednesday 13 January 2010 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Pink or Green)	Nil

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 to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.
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 for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions on this paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N35711A

1. A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.

(*a*) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. State your probabilities clearly.

(3)

(*b*) Find the probability that a blue bead and a green bead are drawn from the jar.

(2)

---

2. The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2|6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

(*a*) the median score,

(1)

(*b*) the interquartile range.

(3)

The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

(*c*) Explain why there is only one employee who will undergo retraining.

(2)

(*d*) Draw a box plot to illustrate the employees' scores.

(3)

---

3. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, x kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum fx = 4841$ and $\sum fx^2 = 15\,889.5$]

(*a*) Write down the missing midpoints in the table above.

(2)

(*b*) Calculate an estimate of the mean birth weight.

(2)

(*c*) Calculate an estimate of the standard deviation of the birth weight.

(3)

(*d*) Use interpolation to estimate the median birth weight.

(2)

(*e*) Describe the skewness of the distribution. Give a reason for your answer.

(2)

4. There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,
70 take developing software,
81 take networking,
35 take developing software and systems support,
28 take networking and developing software,
40 take systems support and networking,
4 take all three extra options.

(*a*) Draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.

Find the probability that the student takes

(*b*) none of the three extra options,

(1)

(*c*) networking only.

(1)

Students who want to become technicians take systems support and networking. Given that a randomly chosen student wants to become a technician,

(*d*) find the probability that this student takes all three extra options.

(2)

---

5. The probability function of a discrete random variable X is given by

$$\mathrm{p}(x) = kx^2, \qquad x = 1, 2, 3.$$

where k is a positive constant.

(*a*) Show that $k = \frac{1}{14}$.

(2)

Find

(*b*) $\mathrm{P}(X \geq 2)$,

(2)

(*c*) $\mathrm{E}(X)$,

(2)

(*d*) $\mathrm{Var}(1 - X)$.

(4)

---

6. The blood pressures, p mmHg, and the ages, t years, of 7 hospital patients are shown in the table below.

Patient	A	B	C	D	E	F	G
t	42	74	48	35	56	26	60
P	98	130	120	88	182	80	135

[$\sum t = 341$, $\sum p = 833$, $\sum t^2 = 18\,181$, $\sum p^2 = 106\,397$, $\sum tp = 42\,948$]

(*a*) Find S_{pp}, S_{tp} and S_{tt} for these data.

(4)

(*b*) Calculate the product moment correlation coefficient for these data.

(3)

(*c*) Interpret the correlation coefficient.

(1)

(*d*) Draw the scatter diagram of blood pressure against age for these 7 patients.

(2)

(*e*) Find the equation of the regression line of p on t.

(4)

(*f*) Plot your regression line on your scatter diagram.

(2)

(*g*) Use your regression line to estimate the blood pressure of a 40 year old patient.

(2)

---

7. The heights of a population of women are normally distributed with mean μ cm and standard deviation σ cm. It is known that 30% of the women are taller than 172 cm and 5% are shorter than 154 cm.

(*a*) Sketch a diagram to show the distribution of heights represented by this information.

(3)

(*b*) Show that $\mu = 154 + 1.6449\sigma$.

(3)

(*c*) Obtain a second equation and hence find the value of μ and the value of σ.

(4)

A woman is chosen at random from the population.

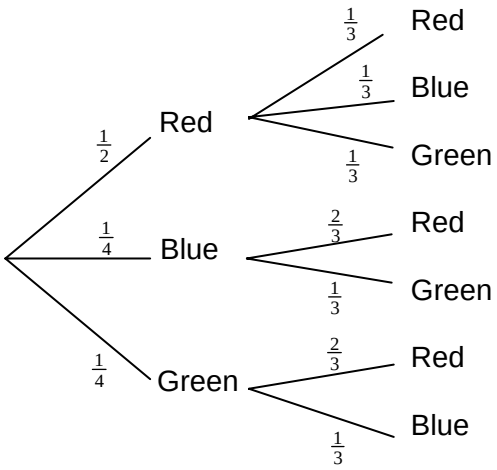
(*d*) Find the probability that she is taller than 160 cm.

(3)

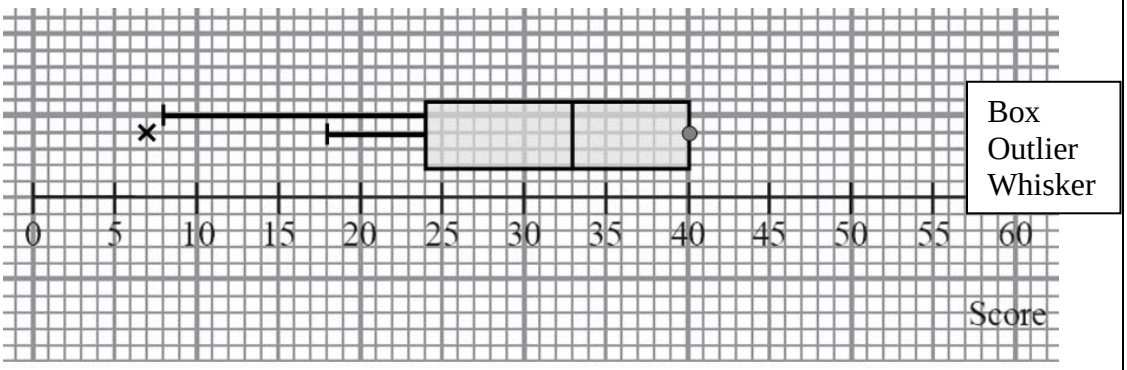
TOTAL FOR PAPER: 75 MARKS

END

January 2010
6683 Statistics S1
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)		M1 A1 A1 (3)
(b)	P(Blue bead and a green bead) $=\left(\frac{1}{4}\times\frac{1}{3}\right)+\left(\frac{1}{4}\times\frac{1}{3}\right)=\frac{1}{6}$ (or any exact equivalent)	M1 A1 (2) **Total [5]**

Q1 (a)	M1 for shape and labels: 3 branches followed by 3,2,2 with some *R*, *B* and *G* seen Allow 3 branches followed by 3, 3, 3 if 0 probabilities are seen implying that 3, 2, 2 intended Allow blank branches if the other probabilities imply probability on blanks is zero Ignore further sets of branches 1st A1 for correct probabilities and correct labels on 1st set of branches. 2nd A1 for correct probabilities and correct labels on 2nd set of branches. (accept 0.33, 0.67 etc or better here)
(b)	M1 for identifying the 2 cases *BG* and *GB* and adding 2 products of probabilities. These cases may be identified by their probabilities e.g. $\left(\frac{1}{4}\times\frac{1}{3}\right)+\left(\frac{1}{4}\times\frac{1}{3}\right)$ NB $\frac{1}{6}$ (or exact equivalent) with no working scores 2/2
Special Case	<u>With Replacement</u> (This oversimplifies so do not apply Mis-Read: max mark 2/5) (a) B1 for 3 branches followed by 3, 3, 3 with correct labels and probabilities of $\frac{1}{2},\frac{1}{4},\frac{1}{4}$ on each. (b) M1 for identifying 2, possibly correct cases and adding 2 products of probabilities but A0 for wrong answer $\left[\left(\frac{1}{4}\times\frac{1}{4}\right)+\left(\frac{1}{4}\times\frac{1}{4}\right)\right]$ will be sufficient for M1A0 here but $\frac{1}{4}\times\frac{1}{2}+\ldots$ would score M0

Question Number		Scheme	Marks
Q2	(a)	Median is 33	B1 (1)
	(b)	$Q_1 = 24, Q_3 = 40, \text{IQR} = 16$	B1 B1 B1ft (3)
	(c)	$Q_1 - \text{IQR} = 24 - 16 = 8$ So 7 is only outlier	M1 A1ft (2)
	(d)	 (accept either whisker)	B1ft B1 B1ft (3) **Total [9]**

Q2 (b) 1st B1 for Q_1 = 24 and 2nd B1 for Q_3 = 40

3rd B1ft for their IQR based on their lower and upper quartile.
Calculation of range (40 – 7 = 33) is B0B0B0

Answer only of IQR = 16 scores 3/3. For any other answer we must see working in (b) or on stem and leaf diagram

(c) M1 for evidence that Q_1-IQR has been attempted, their "8" (>7) seen or clearly attempted is sufficient

A1 ft must have seen their "8" and a suitable comment that only one person scored below this.

(d) 1st B1ft for a clear box shape and ft their Q_1, Q_2 and Q_3 readable off the scale.
Allow this mark for a box shape even if Q_3 = 40, $Q_1 = 7$ and $Q_2 = 33$ are used

2nd B1 for only one outlier appropriately marked at 7

3rd B1ft for either lower whisker. If they choose the whisker to their lower limit for outliers then follow through their "8".
(There should be no upper whisker unless their $Q_3 < 40$, in which case there should be a whisker to 40)

A typical error in (d) is to draw the lower whisker to 7, this can only score B1B0B0

Question Number		Scheme	Marks
Q3	(a)	2.75 or $2\frac{3}{4}$, 5.5 or 5.50 or $5\frac{1}{2}$	B1 B1 (2)
	(b)	Mean birth weight = $\frac{4841}{1500} = 3.227\dot{3}$ **awrt 3.23**	M1 A1 (2)
	(c)	Standard deviation = $\sqrt{\frac{15889.5}{1500} - \left(\frac{4841}{1500}\right)^2} = 0.421093...$ or $s = 0.4212337...$	M1 A1ft A1 (3)
	(d)	$Q_2 = 3.00 + \frac{403}{820} \times 0.5 = 3.2457...$ (allow 403.5….. $\rightarrow$ 3.25)	M1 A1 (2)
	(e)	Mean(3.23)<Median(3.25) (or very close) Negative Skew (or symmetrical)	B1ft dB1ft (2)
			Total [11]

Q3 (b) M1 for a correct expression for mean. Answer only scores both.

(c) M1 for a correct expression (ft their mean) for sd or variance. Condone mis-labelling eg sd=… with no square root or no labelling

1st A1ft for a correct expression (ft their mean) including square root and no mis-labelling
Allow 1st A1 for $\sigma^2 = 0.177... \rightarrow \sigma = 0.42...$

2nd A1 for awrt 0.421. Answer only scores 3/3

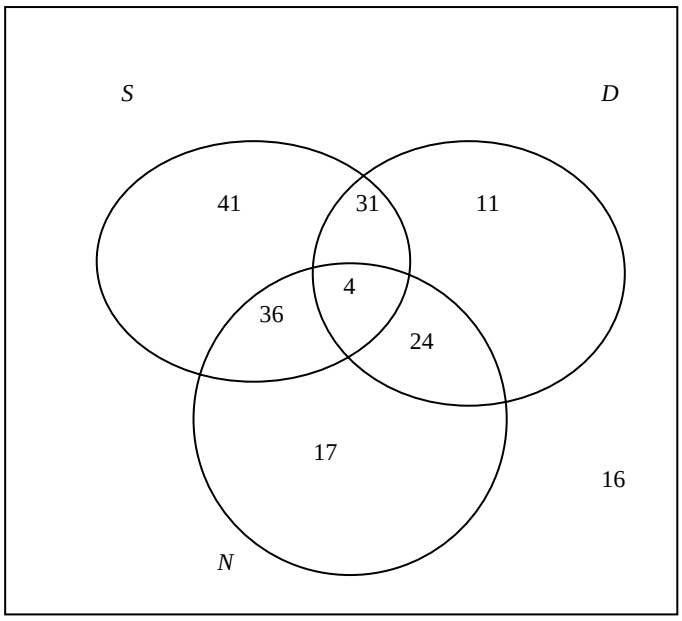
(d) M1 for a correct expression (allow 403.5 i.e. use of $n + 1$) but must have 3.00, 820 and 0.5

A1 for awrt 3.25 provided M1 is scored.
NB 3.25 with no working scores 0/2 as some candidates think mode is 3.25.

(e) 1st B1ft for a comparison of their mean and median (may be in a formula but if ±(mean - median) is calculated that's OK. We are not checking the value but the sign must be consistent.)
Also allow for use of quartiles provided correct values seen: $Q_1 = 3.02, Q_3 = 3.47$

[They should get $(0.22 =) Q_3 - Q_2 < Q_2 - Q_1 (= 0.23)$ and say (slight) negative skew or symmetric]

2nd dB1ft for a compatible comment based on their comparison. Dependent upon a suitable, correct comparison. Mention of "correlation" rather than "skewness" loses this mark.

Question Number	Scheme	Marks
4 (a)		
	3 closed curves and 4 in centre	M1
	Evidence of subtraction	M1
	31,36,24	A1
	41,17,11	A1
	Labels on loops, 16 and box	B1
		(5)
(b)	P(None of the 3 options)= $\frac{16}{180}=\frac{4}{45}$	B1ft
		(1)
(c)	P(Networking only)= $\frac{17}{180}$	B1ft
		(1)
(d)	P(All 3 options/technician)= $\frac{4}{40}=\frac{1}{10}$	M1 A1
		(2)
		Total [9]

4 (a)	2nd M1 There may be evidence of subtraction in "outer" portions, so with 4 in the centre then 35, 40 28 (instead of 31,36,24) along with 33, 9, 3 can score this mark but A0A0 N.B. This is a common error and their "16" becomes 28 but still scores B0 in part (a)
(b)	B1ft for $\frac{16}{180}$ or any exact equivalent. Can ft their "16" from their box. If there is no value for their "16" in the box only allow this mark if they have <u>shown</u> some working.
(c)	B1ft ft their "17". Accept any exact equivalent
(d)	If a probability greater than 1 is found in part (d) score M0A0 M1 for clear sight of $\frac{\mathrm{P}(S\cap D\cap N)}{\mathrm{P}(S\cap N)}$ and an attempt at one of the probabilities, ft their values. Allow P(all 3 | $S\cap N$) = $\frac{4}{36}$ or $\frac{1}{9}$ to score M1 A0. Allow a correct ft from their diagram to score M1A0 e.g. in 33,3,9 case in (a): $\frac{4}{44}$ or $\frac{1}{11}$ is M1A0 A ratio of probabilities with a <u>product</u> of probabilities on top is M0, even with a correct formula. A1 for $\frac{4}{40}$ or $\frac{1}{10}$ or an exact equivalent Allow $\frac{4}{40}$ or $\frac{1}{10}$ to score both marks if this follows from their diagram, otherwise some explanation (method) is required.

Question Number	Scheme	Marks
Q5 (a)	$k + 4k + 9k = 1$ $14k = 1$ $k = \frac{1}{14}$ **given** cso	M1 A1 (2)
(b)	$P(X \geq 2) = 1 - P(X = 1)$ or $P(X = 2) + P(X = 3)$ $= 1 - k = \frac{13}{14}$ or 0.92857... **awrt 0.929**	M1 A1 (2)
(c)	$E(X) = 1 \times k + 2 \times k \times 4 + 3 \times k \times 9$ or $36k$ $= \frac{36}{14} = \frac{18}{7}$ or $2\frac{4}{7}$ (or exact equivalent)	M1 A1 (2)
(d)	$\text{Var}(X) = 1 \times k + 4 \times k \times 4 + 9 \times k \times 9, -\left(\frac{18}{7}\right)^2$ $\text{Var}(1 - X) = \text{Var}(X)$ $= \frac{19}{49}$ or 0.387755... **awrt 0.388**	M1 M1 M1 A1 (4) **Total [10]**

Q5 (a)

M1 for clear attempt to use $\sum p(x) = 1$, full expression needed and the "1" must be clearly seen. This may be seen in a table.

A1cso for no incorrect working seen. The sum and "= 1" must be explicitly seen somewhere.

A verification approach to (a) must show addition for M1 and have a suitable comment e.g. "therefore $k = \frac{1}{14}$" for A1 cso

(b)

M1 for 1- P($X \leq 1$) or P($X = 2$) + P($X = 3$)

A1 for awrt 0.929. Answer only scores 2/2

(c)

M1 for a full expression for E(X) with at least two terms correct.
NB If there is evidence of division (usually by 3) then score M0

A1 for any exact equivalent - answer only scores 2/2

(d)

1st M1 for clear attempt at $E(X^2)$, need at least 2 terms correct in $1 \times k + 4 \times 4k + 9 \times 9k$ or $E(X^2) = 7$

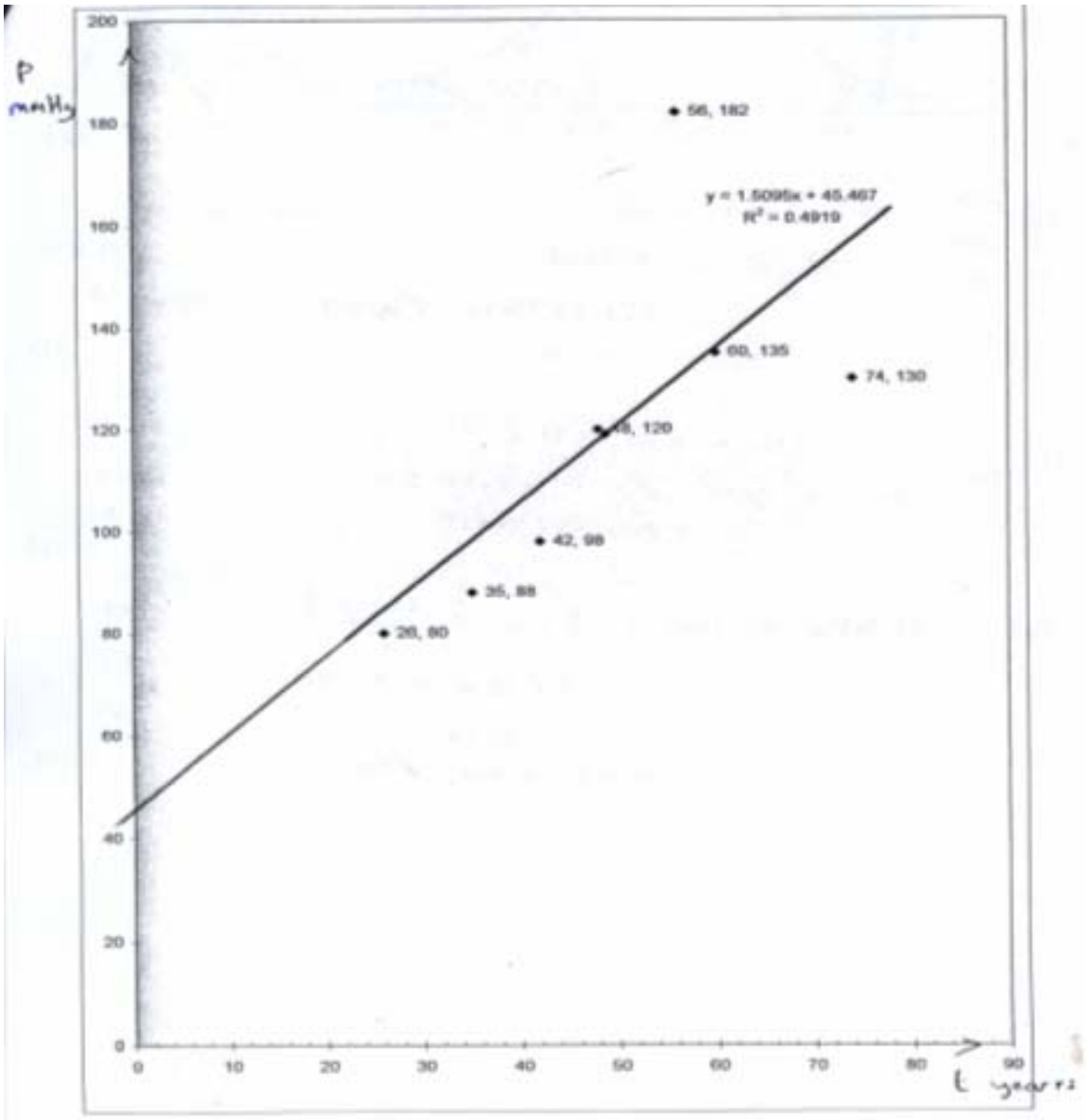
2nd M1 for their $E(X^2) - (\text{their } \mu)^2$

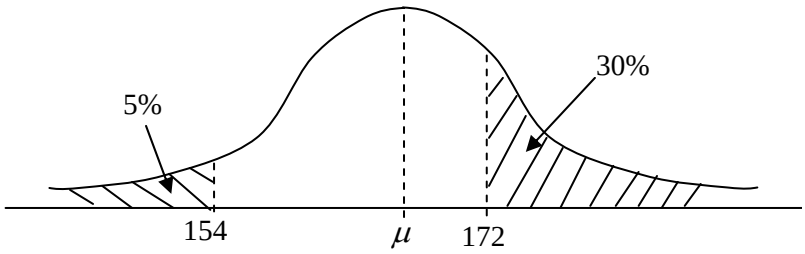
3rd M1 for clearly stating that Var(1 - X) = Var(X), wherever seen

A1 accept awrt 0.388. All 3 M marks are required.
Allow 4/4 for correct answer only but must be for Var(1 – X).

Question Number		Scheme	Marks
Q6	(a)	$S_{pp} = 106397 - \frac{833^2}{7} = 7270$	M1 A1
		$S_{tp} = 42948 - \frac{341 \times 833}{7} = 2369$, $S_{tt} = 18181 - \frac{341^2}{7} = 1569.42857...$ or $\frac{10986}{7}$	A1 A1 (4)
	(b)	$r = \frac{2369}{\sqrt{7270 \times 1569.42857..}}$	M1 A1ft
		$= 0.7013375$ **awrt (0.701)**	A1 (3)
	(c)	(Pmcc shows positive correlation.) Older patients have higher blood pressure	B1 (1)
	(d) + (f)	(d) Points plotted correctly on graph: -1 each error or omission (within one square of correct position) * see diagram below for correct points	B2
		(f) Line drawn with correct intercept, and gradient	B1ft B1 (2+2)
	(e)	$b = \frac{2369}{1569.42857..} = 1.509466...$	M1 A1
		$a = \frac{833}{7} - b \times \frac{341}{7} = 45.467413...$	M1
		$p = 45.5 + 1.51t$	A1 (4)
	(g)	$t = 40$, $p = 105.84...$ from equation or graph. **awrt 106**	M1 A1 (2)
			Total [18]

Q6	(a)	M1 for at least one correct expression 1st A1 for $S_{pp} = 7270$, 2nd A1 for $S_{tp} = 2369$ or 2370 , 3rd A1 for $S_{tt} =$ awrt 1570
	(b)	M1 for attempt at correct formula and at least one correct value (or correct ft) M0 for $\frac{42948}{\sqrt{106397 \times 18181}}$ A1ft All values correct or correct ft. Allow for an answer of 0.7 or 0.70 Answer only: awrt 0.701 is 3/3, answer of 0.7 or 0.70 is 2/3
	(c)	B1 for comment in context that interprets the fact that correlation is positive, as in scheme. Must mention age and blood pressure in words, not just "*t*" and "*p*".
	(d)	Record 1 point incorrect as B1B0 on epen. [NB overlay for (60, 135) is slightly wrong]
	(e)	1st M1 for use of the correct formula for *b*, ft their values from (a) 1st A1 allow 1.5 or better 2nd M1 for use of $\bar{y} - b\bar{x}$ with their values 2nd A1 for full equation with *a* = awrt 45.5 and *b* = awrt 1.51. Must be *p* in terms of *t*, not *x* and *y*.
	(f)	1st B1ft ft their intercept (within one square). You may have to extend their line. 2nd B1 for correct gradient i.e. parallel to given line (Allow 1 square out when *t* = 80)
	(g)	M1 for clear use of their equation with *t* = 40 or correct value from their graph. A1 for awrt 106. Correct answer only (2/2) otherwise look for evidence on graph to award M1

Question Number	Scheme
Q6 (d) + (f)	Diagram for Q6 (d) + (f) 

Question Number	Scheme	Marks
Q7 (a)	 bell shaped, must have inflexions 154,172 on axis 5% and 30%	B1 B1 B1 (3)
(b)	$P(X < 154) = 0.05$ $\frac{154-\mu}{\sigma} = -1.6449$ or $\frac{\mu-154}{\sigma} = 1.6449$ $\mu = 154 + 1.6449\sigma$ **given**	M1 B1 A1 cso (3)
(c)	$172 - \mu = 0.5244\sigma$ or $\frac{172-\mu}{\sigma} = 0.5244$ (allow $z = 0.52$ or better here but must be in an equation) Solving gives $\sigma = 8.2976075$ (**awrt 8.30)** and $\mu = 167.64873$ (**awrt 168**)	B1 M1 A1 A1 (4)
(d)	$P(\text{Taller than 160cm}) = P\left(Z > \frac{160-\mu}{\sigma}\right)$ $= P(Z < 0.9217994)$ $= 0.8212$ **awrt 0.82**	M1 B1 A1 (3) **Total [13]**

(a) 2nd B1 for 154 and 172 marked but 154 must be < μ and 172 > μ. But μ need not be marked.
Allow for $\frac{154-\mu}{\sigma}$ and $\frac{172-\mu}{\sigma}$ marked on appropriate sides of the peak.
3rd B1 the 5% and 30% should be clearly indicated in the correct regions i.e. LH tail and RH tail.

(b) M1 for $\pm\frac{(154-\mu)}{\sigma} = z$ value (z must be recognizable e.g. 1.64, 1.65, 1.96 but NOT 0.5199 etc)
B1 for $\pm$ 1.6449 seen in a line before the final answer.
A1cso for no incorrect statements (in μ, σ) equating a z value and a probability or incorrect signs
e.g. $\frac{154-\mu}{\sigma} = 0.05$ or $\frac{154-\mu}{\sigma} = 1.6449$ or $P(Z < \frac{\mu-154}{\sigma}) = 1.6449$

(c) B1 for a correct 2nd equation (NB $172 - \mu = 0.525\sigma$ is B0, since z is incorrect)
M1 for solving their two linear equations leading to $\mu = ...$ or $\sigma = ...$
1st A1 for σ = awrt 8.30, 2nd A1 for μ = awrt 168 [NB the 168 can come from false working.
These A marks require use of correct equation from (b), and a z value for "0.5244" in (c)]
NB use of $z = 0.52$ will typically get σ=8.31 and μ = 167.67… and score B1M1A0A1
No working and both correct scores 4/4, only one correct scores 0/4
Provided the M1 is scored the A1s can be scored even with B0 (e.g. for z =0.525)

(d) M1 for attempt to standardise with 160, their μ and their σ (> 0). Even allow with symbols μ and σ.
B1 for z = awrt $\pm$ 0.92
No working and a correct answer can score 3/3 provided σ and μ are correct to 2sf.