Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Level

Friday 14 January 2011 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	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 formulas 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 S2), the paper reference (6684), 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.
This paper has 8 questions.
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 not gain full credit.

1. A random sample of 50 salmon was caught by a scientist. He recorded the length l cm and weight w kg of each salmon.

The following summary statistics were calculated from these data.

$\sum l = 4027$ $\sum l^2 = 327\,754.5$ $\sum w = 357.1$ $\sum lw = 29\,330.5$ $S_{ww} = 289.6$

(*a*) Find S_{ll} and S_{lw}.

(3)

(*b*) Calculate, to 3 significant figures, the product moment correlation coefficient between l and w.

(2)

(*c*) Give an interpretation of your coefficient.

(1)

---

2. Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$\sum x = 84.6$

(*a*) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0.

Keith corrects these figures.

(*b*) State, giving your reason, the effect this will have on the mean.

(2)

---

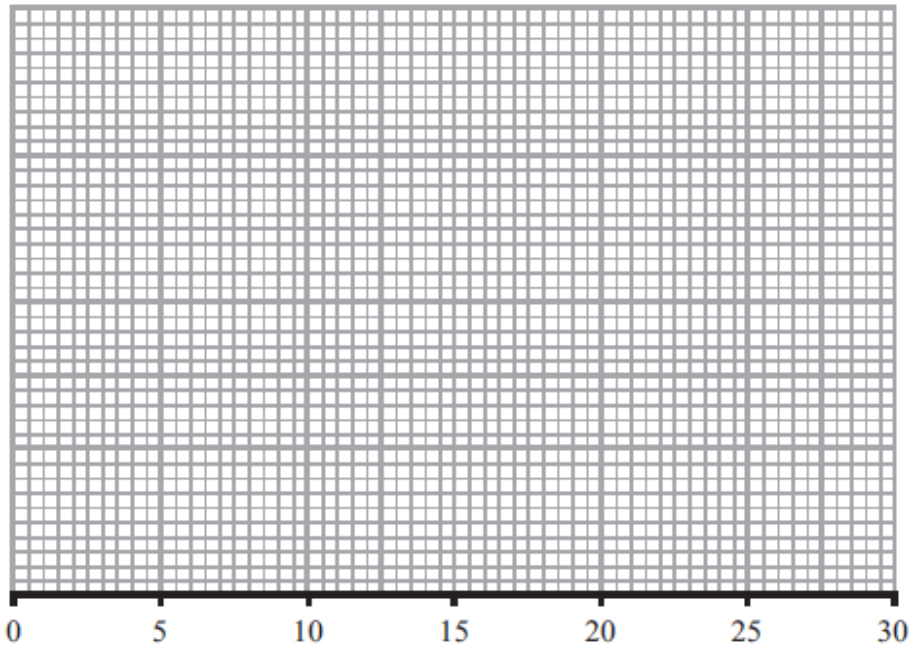
3. Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls
either 1.5 × interquartile range above the upper quartile
or 1.5 × interquartile range below the lower quartile.

(*a*) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers.

(5)



(*b*) State the skewness of the distribution of the amount of sales received. Justify your answer.

(2)

(*c*) The company claims that for 75 % of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer.

(2)

4. A farmer collected data on the annual rainfall, x cm, and the annual yield of peas, p tonnes per acre.

The data for annual rainfall was coded using $v = \frac{x-5}{10}$ and the following statistics were found.

$$S_{vv} = 5.753 \qquad S_{pv} = 1.688 \qquad S_{pp} = 1.168 \qquad \bar{p} = 3.22 \qquad \bar{v} = 4.42$$

(*a*) Find the equation of the regression line of p on v in the form $p = a + bv$.

(4)

(*b*) Using your regression line estimate the annual yield of peas per acre when the annual rainfall is 85 cm.

(2)

---

5. On a randomly chosen day, each of the 32 students in a class recorded the time, t minutes to the nearest minute, they spent on their homework. The data for the class is summarised in the following table.

Time, t	Number of students
10 – 19	2
20 – 29	4
30 – 39	8
40 – 49	11
50 – 69	5
70 – 79	2

(*a*) Use interpolation to estimate the value of the median.

(2)

Given that

$$\sum t = 1414 \qquad \text{and} \qquad \sum t^2 = 69\,378,$$

(*b*) find the mean and the standard deviation of the times spent by the students on their homework.

(3)

(*c*) Comment on the skewness of the distribution of the times spent by the students on their homework. Give a reason for your answer.

(2)

---

6. The discrete random variable X has the probability distribution

x	1	2	3	4
$P(X = x)$	k	$2k$	$3k$	$4k$

(*a*) Show that $k = 0.1$

(1)

Find

(*b*) $E(X)$

(2)

(*c*) $E(X^2)$

(2)

(*d*) $\text{Var}(2 - 5X)$

(3)

Two independent observations X_1 and X_2 are made of X.

(*e*) Show that $P(X_1 + X_2 = 4) = 0.1$

(2)

(*f*) Complete the probability distribution table for $X_1 + X_2$.

(2)

y	2	3	4	5	6	7	8
$P(X_1 + X_2 = y)$	0.01	0.04	0.10		0.25	0.24	

(g) Find $P(1.5 < X_1 + X_2 \leq 3.5)$

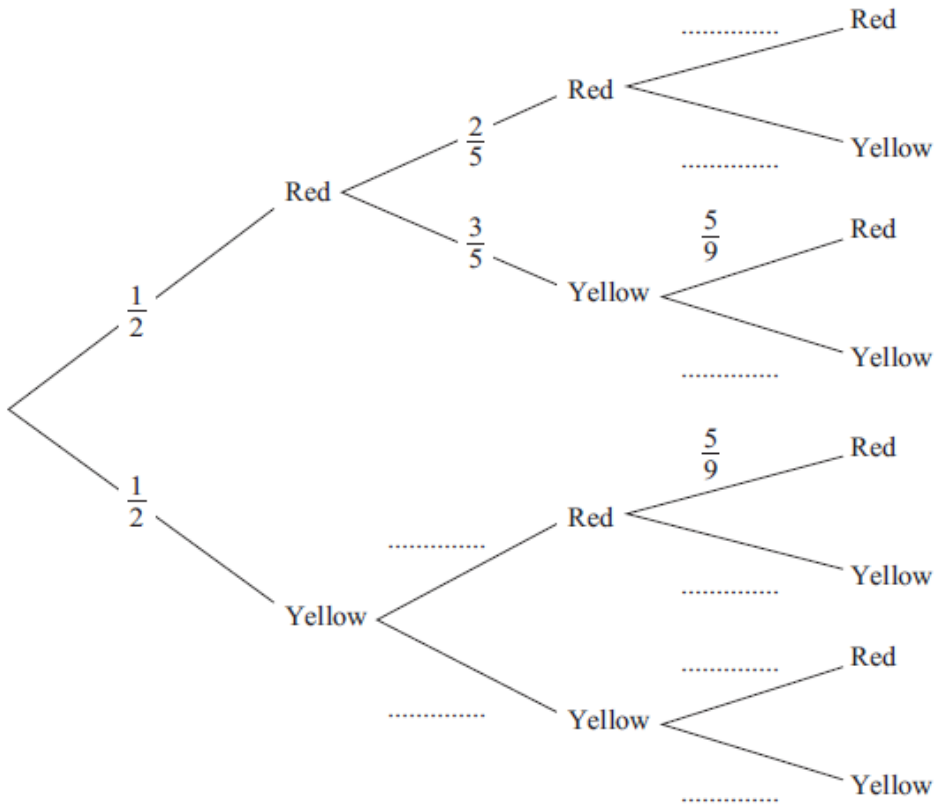
(2)

---

7. The bag P contains 6 balls of which 3 are red and 3 are yellow.
The bag Q contains 7 balls of which 4 are red and 3 are yellow.
A ball is drawn at random from bag P and placed in bag Q. A second ball is drawn at random from bag P and placed in bag Q.
A third ball is then drawn at random from the 9 balls in bag Q.

The event A occurs when the 2 balls drawn from bag P are of the same colour.
The event B occurs when the ball drawn from bag Q is red.

(a) Copy and complete the tree diagram shown below.



(4)

(b) Find P(A).

(3)

(c) Show that P(B) = $\frac{5}{9}$.

(3)

(d) Show that P($A \cap B$) = $\frac{2}{9}$.

(2)

(e) Hence find P($A \cup B$).

(2)

(f) Given that all three balls drawn are the same colour, find the probability that they are all red.

(3)

8. The weight, X grams, of soup put in a tin by machine A is normally distributed with a mean of 160 g and a standard deviation of 5 g.

A tin is selected at random.

(a) Find the probability that this tin contains more than 168 g.

(3)

The weight stated on the tin is w grams.

(b) Find w such that $P(X < w) = 0.01$.

(3)

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.

(c) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$, find the value of μ and the value of σ.

(6)

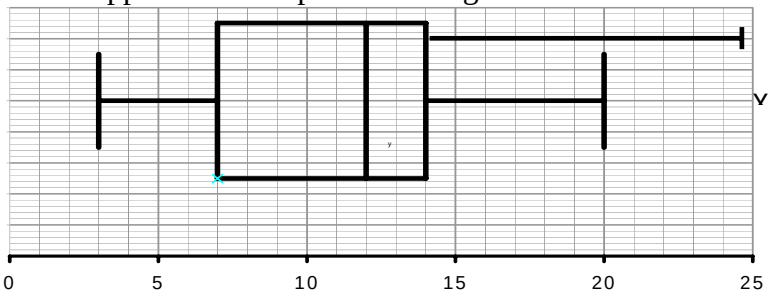
TOTAL FOR PAPER: 75 MARKS

END

January 2011
Statistics S1 6683
Mark Scheme

Question Number	Scheme	Marks
1. **(a)**	$S_{ll} = 327754.5 - \frac{4027^2}{50} = 3419.92$ $S_{lw} = 29330.5 - \frac{357.1 \times 4027}{50} = 569.666$	M1 A1 A1 (3)
(b)	$r = \frac{569.666}{\sqrt{3419.92 \times 289.6}} = 0.572$ awrt 0.572 or 0.573	M1 A1 (2)
(c)	As the length of the salmon increases the weight increases	B1ft (1) **[6]**
	Notes	
(a)	M1 for at least one correct expression 1st A1 for S_{ll} = awrt 3420 (Condone S_{xx} =… or even S_{yy} =…) 2nd A1 for S_{lw} = awrt 570 (Condone S_{xy} =…)	
(b)	M1 for attempt at correct formula. Must have their S_{ll}, S_{lw} and given S_{ww} in the correct places If S_{ll}, S_{lw} are correct and an answer of awrt 0.57 is seen then award M1A0 M0 for $\frac{29330.5}{\sqrt{327754.5 \times 289.6}}$	
(c)	B1ft for a comment mentioning "length" and "weight", not just *l* and *w*, and the idea of longer salmon weighing more. e.g. "positive correlation between weight and length" is B0 since the idea of positive correlation is not explained. Allow "larger" instead of "heavier" or "longer" Ignore any spurious values mentioned such as 0.572 If their *r* is negative (but must be $r > -1$) ft an appropriate comment. Condone $r > 1$ if comment is correct. If $\lvert r \rvert < 0.4$ allow a comment of no or little relationship between weight and length but for $0 < r < 0.4$ the printed answer is still acceptable too. Treat mention of "skewness" as ISW if a correct interpretation is given	

Question Number	Scheme	Marks
2. **(a)**	$2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 + 84.6 = 107$ mean = 107 / 28 (= 3.821…) (**awrt 3.8)**	M1 A1 (2)
(b)	It will have no effect since one is 4.5 under what it should be and the other is 4.5 above what it should be.	B1 dB1 (2) **[4]**
	Notes	
(a)	M1 for a clear attempt to add the two sums. Accept a full expression or $2.8 + 5.6 + \ldots + 84.6 = x$ where $100<x<110$ i.e. seeing at least two correct terms of Keith's and the 84.6 with a slip. A1 for awrt 3.8 (Condone 1 dp/2sf here since data is given to 1 dp or 2 sf) Accept $\frac{107}{28}$ or $3\frac{23}{28}$ or any exact equivalent **Correct answer implies M1A1**	
(b)	1st B1 for clearly stating that it will have no effect. ("roughly the same" is B0 B0) 2nd dB1 for a supporting reason that mentions the fact that the increase and decrease are the same and gives some numerical value(s) to support this. e.g. Sum of Keith's observations is still 22.4 (or mean is still 3.2) or Sum is still 107 or $9.4 - 4.9 = 5 - 0.5$ (o.e.) This second B1 is dependent on their saying there is no effect so B0B1 is not possible.	

Question Number	Scheme	Marks
3. **(a)**	Outliers $14 + 1.5\times(14-7) = 24.5$ $7 - 1.5\times(14-7) = -3.5$ Outlier 25 either upper limit acceptable on diagram  Sales in £'000	M1 A1 M1 A1ft B1 (5)
(b)	Since $Q_3 - Q_2 < Q_2 - Q_1$. Allow written explanation negatively skew	B1 dB1 (2)
(c)	not true since the lower quartile is 7000 and therefore 75% above 7000 not 10000 or 10 is inside the box or any other sensible comment	B1 dB1 (2) **[9]**

Notes

(a)

A fully correct box-plot (either version) with no supporting work scores 5/5. Otherwise read on

1st M1 for at least one correct calculation seen

1st A1 for 24.5 and −3.5 (or just negative noted) seen. May be read off the graph.
If both values are seen but no calculation is given then M1A1,one value M1A0.

2nd M1 for a box with an upper and a lower whisker(s) with at least 2 correct values
(condone no median marked)

2nd A1ft for 3,7, 12, 14 and 20 or 24.5 in appropriate places and readable off their scale
If <u>both</u> upper whiskers are seen A0
Apply ft for their <u>whiskers</u> being compatible with their <u>outlier limits</u>
e.g. if their lower limit is + 3.5 then a lower whisker ending at 4 or 3.5 is OK

B1 for only one outlier appropriately marked at 25

Apply <u>+</u> 0.5 square accuracy for diagram

(b)

1st B1 for $Q_3 - Q_2 < Q_2 - Q_1$ statement or an equivalent statement in words
Use of $Q_3 - Q_2 < Q_2 - Q_1$ does not require differences to be seen.

2nd dB1 for "negative skew" dependent on suitable reason given above. "correlation" is B0
"positive skew" with a supporting argument based on whiskers can score B1B1
e.g. "right hand whisker is longer than LH one so positive skew"
$Q_3 - Q_2 < Q_2 - Q_1$ followed by "positive skew" is B1B0

(c)

1st B1 for rejecting the company's claim

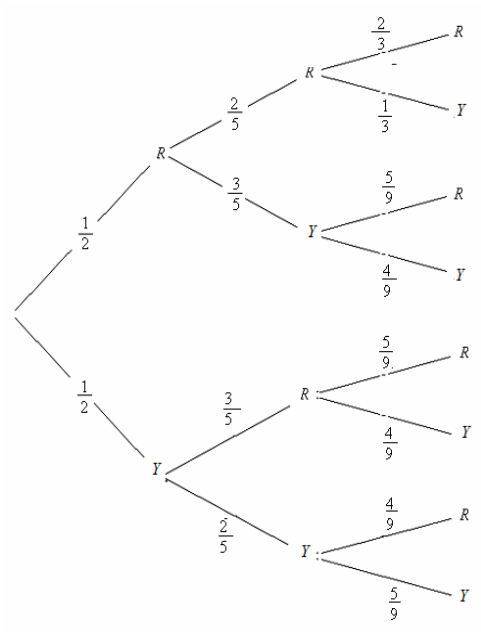
2nd dB1 for an appropriate supporting reason. Dependent on rejecting company's claim.

Question Number	Scheme	Marks
4. **(a)**	$b = \frac{1.688}{5.753} = 0.293$ $a = 3.22 - 4.42 \times 0.293 = 1.9231...$ $p = 1.92 + 0.293v$	M1A1 M1 A1 (4)
(b)	$v = \frac{85-5}{10} = 8$ $p = 1.92 + 0.293 \times 8 = 4.3$ **(awrt 4.3)**	M1 A1 (2) **[6]**
	Notes	
(a)	**Can ignore (a) and (b) labels here** 1st M1 for a correct expression for b. $\frac{1.688}{1.168}$ is M0 1st A1 for awrt 0.29 2nd M1 for use of $a = \bar{p} - b\bar{v}$ follow through their value of b(or even just the letter b) 2nd A1 for a complete equation with a = awrt 1.92 and b = awrt 0.293 y or p = 1.92 + 0.293x is A0 Correct answer with no working is 4/4	
(b)	M1 for an attempt to find the value of v when x = 85 (at least 2 correct terms in $\pm\frac{85-5}{10}$) or for an attempt to find an equation for p in terms of x and using x = 85 Attempt at equation of p in x requires $p = 1.92 + 0.293\frac{(x-5)}{10}$ A1 for awrt 4.3 (award when first seen and apply ISW) N.B. $p = 1.92 + 0.293 \times 85$ (o.e.) is M0A0	

Question Number	Scheme	Marks
5. **(a)**	Median = 32/2 = 16th term (16.5) $\frac{x-39.5}{49.5-39.5}=\frac{16-14}{25-14}$ or $x=39.5+\left(\frac{2}{11}\times 10\right)$ Median = 41.3 (use of n + 1 gives 41.8) **(awrt 41.3)**	M1 A1 (2)
(b)	Mean$=\frac{1414}{32}=44.1875$ **(awrt 44.2)** Standard deviation $=\sqrt{\frac{69378}{32}-\left(\frac{1414}{32}\right)^2}$ = 14.7 (or s = 14.9)	B1 M1 A1 (3)
(c)	mean > median therefore positive skew	B1ft B1ft (2) **[7]**
	Notes	
(a)	M1 for an attempt to use interpolation to find the median. Condone use of 39 or 40 for 39.5 e.g. allow $39+\frac{2}{11}\times 10$ (o.e.) or $40+\frac{2}{11}\times 10$ (o.e.) to score M1A0 but must have the 10 A1 for awrt 41.3 (or awrt 41.8 if using (n + 1))	
(b) **Mid-points**	B1 for awrt 44.2 M1 for a correct expression including square root. (Allow ft of their mean) A1 for awrt 14.7 (If using s for awrt 14.9) You may see $\sum t=1339\rightarrow \bar{t}=41.8$ and $\sum t^2=62928\rightarrow$ σ 14.7 or s = 14.9 this scores B0 for the mean but can score M1 for a correct st.dev expression and A1 for ans. **Correct answer only in (a) and (b) can score full marks but check (n +1) case in (a)**	
(c) **Quartiles**	1st B1ft for a correct comparison of their mean and their median (may be in a formula) Calculating median – mean as negative is OK for this B1 but must say +ve skew for 2nd B1 Only allow comparison to be ≈ 0 if $\lvert\text{mean} - \text{median}\rvert \le 0.5$ 2nd B1ft for a correct description of skewness based on their values of mean and median. ft their values for mean and median not their previous calculation/comparison Must be compatible with their previous comparison (if they have one) "Positive skew" with no reason is B0B1 provided you can see their values that imply that. Description should be "positive" or "negative" or "no" skew or "symmetric" "Positive correlation" is B0 1st B1ft if Q_1 = awrt 32 and Q_3 = awrt 49 seen and a correct comparison made. ft Q_2 2nd B1ft if Q_1 = awrt 32 or Q_3 = awrt 49 seen and a correct description based on their quartiles and their comparison is made. (Should get "negative skew")	

Question Number	Scheme	Marks
6. **(a)**	$k + 2k + 3k + 4k = 1$ or $10k = 1$ $k = 0.1$ (*****) [allow verification with a comment e.g. "so $k = 0.1$"]	B1cso (1)
(b)	$\mathrm{E}(X) = 1\times0.1 + 2\times0.2 + 3\times0.3 + 4\times0.4 = 3$	M1 A1 (2)
(c)	$\mathrm{E}(X^2) = 1\times0.1 + 4\times0.2 + 9\times0.3 + 16\times0.4 = 10$	M1 A1 (2)
(d)	$\mathrm{Var}(X) = 10 - 9(=1)$ $\mathrm{Var}(2-5X) = 5^2\mathrm{Var}(X) = 25$	M1 M1 A1 (3)
(e)	$\mathrm{P}(1,3) + \mathrm{P}(2,2) = 2\times0.1\times0.3 + 0.2\times0.2 = 0.1$ (*****)	M1 A1cso (2)
(f)	$X_1 + X_2$: 2, 3, 4, 5, 6, 7, 8 p: 0.01, 0.04, 0.1, **0.2**, 0.25, 0.24, **0.16**	B1 B1 (2)
(g)	$\mathrm{P}(2) + \mathrm{P}(3) = 0.05$	M1A1 (2) **[14]**

Question Number	Scheme	Marks
	Notes	
(a)	B1 for a clear attempt to use sum of probabilities = 1. Must see previous line as well as $k = 0.1$ **A correct expression for** $\mathrm{E}(X)$ or $\mathrm{E}(X^2)$ **that is later divided by 4 scores M0**	
(b)	M1 for a completely correct expression. May be implied by correct answer of 3 or $30k$ A1 for 3 only.	
(c)	M1 for a completely correct expression. May be implied by correct answer of 10 or $100k$ A1 for 10 only. [For $\mathrm{E}(X^2) = 0.1 + 0.8 + 2.7 + 6.4 - 9$ = 1scores M0A0 but accept this as Var(X) in (d)]	
(d)	1st M1 for using Var(X) = $\mathrm{E}(X^2) - \mathrm{E}(X)^2$, f.t their values from (b) and (c) Allow this mark for Var(X) = $10 - 9$ or better. May be implied if this is seen in (c). 2nd M1 for $5^2\mathrm{Var}(X)$ or 25Var(X) can f.t. their Var(X). Allow -5^2 if it later becomes +25 A1 for 25 only. Dependent upon both Ms Forming distribution for Y =2-5X gets M1 for E(Y^2)=194 then M1A1 for 194-169=25	
(e)	M1 for correctly identifying (1, 3) or (3, 1) and (2, 2) as required cases ($3k^2 + 4k^2$ or better) A1 cso for 0.1 only but must see evidence for M1	
(f)	1st B1 for 0.2 correctly assigned. May be in table. 2nd B1 for 0.16 correctly assigned. May be in table	
(g)	M1 for P(2) + P(3). May be implied by correct answer of 0.05 A1 for 0.05 only. **Correct answer only can score full marks in parts (b), (c), (f) and (g)**	

Question Number	Scheme	Marks
7. (a)	 $\frac{2}{15}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{4}{45}$ $\left(\frac{1}{9}\right)$ both $\frac{2}{3},\frac{1}{3}$ $\frac{4}{9}$ both $\frac{3}{5},\frac{2}{5}$ all three of $\frac{4}{9},\frac{4}{9},\frac{5}{9}$	B1 B1 B1 B1 (4)
(b)	$P(A) = P(RR) + P(YY) = \frac{1}{2}\times\frac{2}{5} + \frac{1}{2}\times"\frac{2}{5}" = \frac{2}{5}$ B1 for $\frac{1}{2}\times\frac{2}{5}$ (oe) seen at least once	B1 M1 A1 (3)
(c)	P(*B*) = P(*RRR*) + P(*RYR*) + P(*YRR*) + P(*YYR*) M1 for at least 1 case of 3 balls identified. (Implied by 2nd M1)	M1
	$\left(\frac{1}{2}\times\frac{2}{5}\times"\frac{2}{3}"\right)+\left(\frac{1}{2}\times\frac{3}{5}\times\frac{5}{9}\right)+\left(\frac{1}{2}\times"\frac{3}{5}"\times\frac{5}{9}\right)+\left(\frac{1}{2}\times"\frac{2}{5}"\times"\frac{4}{9}"\right) = \underline{\frac{5}{9}}$ **(*)**	M1,A1cso (3)
(d)	$P(A\cap B) = P(RRR) + P(YYR)$ M1 for identifying both cases and + probs. may be implied by correct expressions	M1
	$= \left(\frac{1}{2}\times\frac{2}{5}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{2}{5}\times\frac{4}{9}\right) = \underline{\frac{2}{9}}$ **(*)**	A1cso (2)
(e)	$P(A\cup B) = P(A) + P(B) - P(A\cap B)$ Must have some attempt to use $= "\frac{2}{5}" + \frac{5}{9} - \frac{2}{9} = \underline{\frac{11}{15}}$	M1 A1cao (2)

Question Number	Scheme		Marks
(f)	$\dfrac{\mathrm{P}(RRR)}{\mathrm{P}(RRR)+\mathrm{P}(YYY)} = \dfrac{\frac{1}{2}\times\frac{2}{5}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{5}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{2}{5}\times\frac{5}{9}\right)} = \dfrac{6}{11}$	Probabilities must come from the product of 3 probs. from their tree diagram.	M1 A1ft A1 cao (3) **[17]**
	Notes		
(b)	M1 for both cases, and +, attempted, ft their values from tree diagram. May be 4 cases of 3 balls.		
(c)	2nd M1 for all 4 correct expressions, ft their values from tree diagram. A1 is cso		
(e)	M1 for clear attempt to use the correct formula, must have some correct substitution. ft their (b)		
(f)	M1 for identifying the correct probabilities and forming appropriate fraction of probs. 1st A1ft for a correct expression using probabilities from their tree **Accept exact decimal equivalents. Correct answer only is full marks except in (c) and (d)**		

Question Number	Scheme	Marks
8. **(a)**	$P(X > 168) = P\left(Z > \frac{168-160}{5}\right)$	M1
	$= P(Z > 1.6)$	A1
	$= 0.0548$ **awrt 0.0548**	A1 (3)
(b)	$P(X < w) = P\left(Z < \frac{w-160}{5}\right)$	
	$\frac{w-160}{5} = -2.3263$	M1 B1
	$w = 148.37$ **awrt 148**	A1 (3)
(c)	$\frac{160-\mu}{\sigma} = 2.3263$	M1 B1
	$\frac{152-\mu}{\sigma} = -1.2816$	B1
	$160 - \mu = 2.3263\sigma$	
	$152 - \mu = -1.2816\sigma$	
	$8 = 3.6079\sigma$	M1
	$\sigma = 2.21\ldots$ **awrt 2.22**	A1
	$\mu = 154.84\ldots$ **awrt 155**	A1 (6)
		[12]

Notes

(a)

M1 for an attempt to standardize 168 with 160 and 5 i.e. $\pm\left(\frac{168-160}{5}\right)$ or implied by 1.6

1st A1 for P(Z > 1.6) or P(Z < -1.6) ie z = 1.6 and a correct inequality or 1.6 on a shaded diagram

Correct answer to (a) implies all 3 marks

(b)

M1 for attempting $\pm\left(\frac{w-160}{5}\right)$ = recognizable z value ($|z| > 1$)

B1 for $z = \pm 2.3263$ or better. Should be $z =\ldots$ or implied so: $1 - 2.3263 = \frac{w-160}{5}$ is M0B0

A1 for awrt 148. This may be scored for other z values so M1B0A1 is possible

For awrt 148 only with no working seen award M1B0A1

(c)

M1 for attempting to standardize 160 or 152 with μ and σ (allow $\pm$) and equate to z value ($|z|>1$)

1st B1 for awrt $\pm$ 2.33 or $\pm$ 2.32 seen

2nd B1 for awrt $\pm$ 1.28 seen

2nd M1 for attempt to solve their two linear equations in μ and σ leading to equation in just one variable

1st A1 for σ = awrt 2.22 . Award when 1st seen

2nd A1 for μ = awrt 155. Correct answer only for part (c) can score all 6 marks.

NB σ = 2.21 commonly comes from z = 2.34 and usually scores M1B0B1M1A0A1

The A marks in (c) require both M marks to have been earned