Paper Reference(s)

6683/01
Edexcel GCE
Statistics S1
Silver Level S2

Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question**
Mathematical Formulae (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 formulas stored in them.

Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
There are 7 questions in this question 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.

Suggested grade boundaries for this paper:

A*	A	B	C	D	E
67	60	52	44	36	30

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)

   January 2010

2. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

   [You may use $\sum x^2 = 724\,961$]

   (*a*) Find the mean and standard deviation of the level of cotinine in a patient's blood.

   (4)

   (*b*) Find the median, upper and lower quartiles of these data.

   (3)

   A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3 + 1.5(Q_3 - Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

   (*c*) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.

   (4)

   Research suggests that cotinine levels in the blood form a skewed distribution.

   One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

   (*d*) Evaluate this measure and describe the skewness of these data.

   (3)

   January 2008

3.   The discrete random variable Y has the probability distribution

y	1	2	3	4
$P(Y = y)$	a	b	0.3	c

where a, b and c are constants.

The cumulative distribution function F(y) of Y is given in the following table.

y	1	2	3	4
F(y)	0.1	0.5	d	1.0

where d is a constant.

(*a*) Find the value of a, the value of b, the value of c and the value of d.

(5)

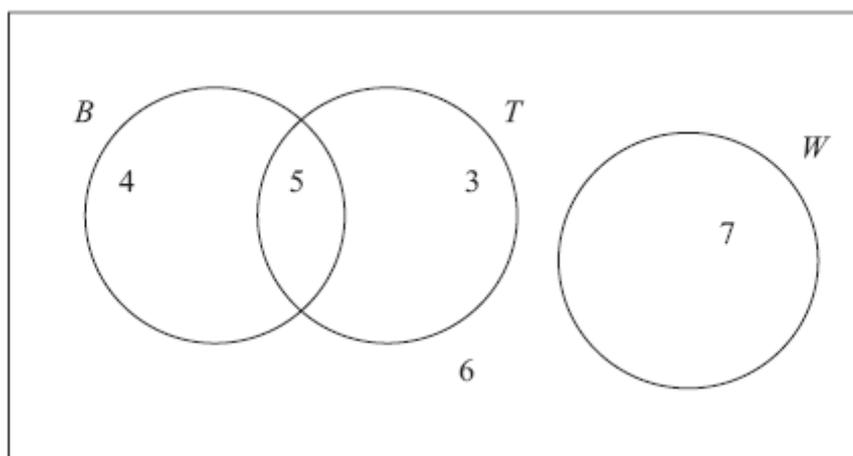(*b*) Find $P(3Y + 2 \geq 8)$.

(2)

May 2011

4.

Figure 1

Figure 1 shows how 25 people travelled to work.

Their travel to work is represented by the events

B bicycle

T train

W walk

(*a*) Write down 2 of these events that are mutually exclusive. Give a reason for your answer.

(2)

(*b*) Determine whether or not B and T are independent events.

(3)

One person is chosen at random.

Find the probability that this person

(*c*) walks to work,

(1)

(*d*) travels to work by bicycle and train.

(1)

Given that this person travels to work by bicycle,

(*e*) find the probability that they will also take the train.

(2)

May 2012

5. The weight, w grams, and the length, l mm, of 10 randomly selected newborn turtles are given in the table below.

l	49.0	52.0	53.0	54.5	54.1	53.4	50.0	51.6	49.5	51.2
w	29	32	34	39	38	35	30	31	29	30

(You may use $S_{ll} = 33.381$ $S_{wl} = 59.99$ $S_{ww} = 120.1$)

(*a*) Find the equation of the regression line of w on l in the form $w = a + bl$.

(5)

(*b*) Use your regression line to estimate the weight of a newborn turtle of length 60 mm.

(2)

(*c*) Comment on the reliability of your estimate giving a reason for your answer.

(2)

May 2009

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)

January 2011

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.
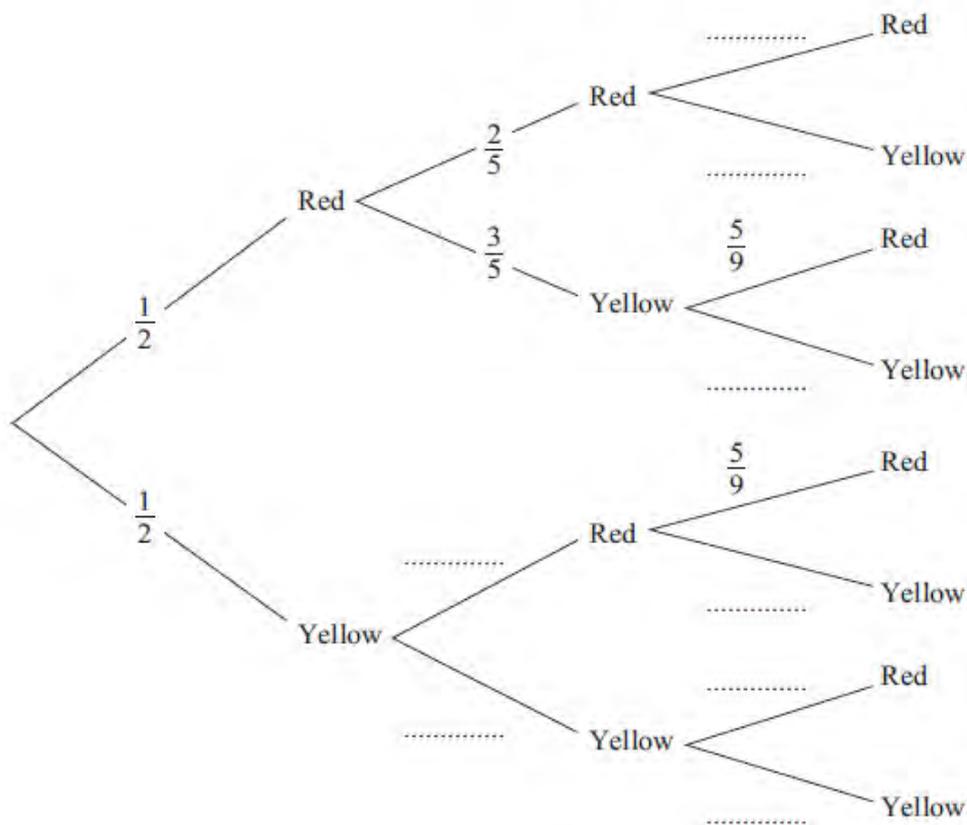
(4)

(*b*) Find P(*A*).

(3)

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

(3)

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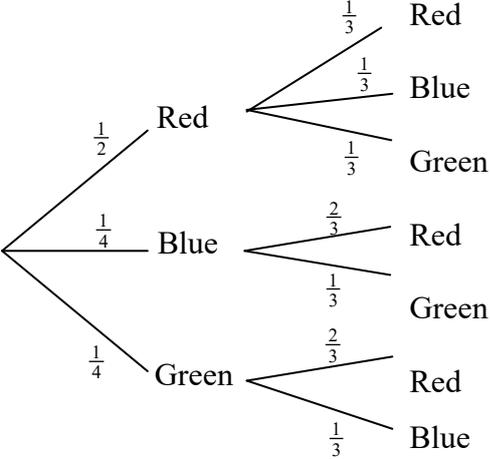
(2)

(*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)

January 2011

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)		M1 A1 A1 (3)
(b)	P(Blue bead and a green bead) $= \left(\dfrac{1}{4} \times \dfrac{1}{3}\right) + \left(\dfrac{1}{4} \times \dfrac{1}{3}\right) = \dfrac{1}{6}$ (o.e.)	M1 A1 (2) **[11]**
2. (a)	mean is $\dfrac{2757}{12}$, $= 229.75$ awrt 230	M1, A1
	sd is $\sqrt{\dfrac{724961}{12} - (229.75)^2}$, $= 87.34045$ awrt 87.3	M1, A1 (4)
(b)	Ordered list is: 125, 160, 169, 171, 175, 186, 210, 243, 250, 258, 390, 420 $Q_2 = \frac{1}{2}(186 + 210) = 198$ $Q_1 = \frac{1}{2}(169 + 171) = 170$ $Q_3 = \frac{1}{2}(250 + 258) = 254$	B1 B1 B1 (3)
(c)	$Q_3 + 1.5(Q_3 - Q_1) = 254 + 1.5(254 - 170), = 380$ Accept awrt (370-392) Patients F (420) and B (390) are outliers.	M1, A1 B1ftB1ft (4)
(d)	$\dfrac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1} = \dfrac{170 - 2 \times 198 + 254}{254 - 170}, = 0.3$ awrt 0.33 Positive skew.	M1, A1 A1ft (3) **[14]**

Question Number	Scheme	Marks
3. (a)	$a = \underline{0.1}$	B1
	$[F(3) = F(2) + P(Y = 3) = (0.5 + 0.3)]$	
	$d = \underline{0.8}$	B1
	$b = F(2) - a = 0.5 - 0.1$ <u>or</u> $a + b = 0.5$	M1
	$b = \underline{0.4}$	A1
	$c = 1 - F(3)$ <u>or</u> $1 - (a + b + 0.3)$ <u>or</u> $a + b + c = 0.7$	
	$c = \underline{0.2}$	A1
		(5)
(b)	$P(3Y + 2 \geq 8) = P(Y \geq 2)$ <u>or</u> $1 - P(Y \leq 1)$	M1
	$= b + 0.3 + c$ <u>or</u> $1 - a$ $= \underline{0.9}$	A1ft
		(2)
		[7]
4. (a)	B, W <u>or</u> T, W [accept $B \cup T, W$ <u>or</u> $B \cap T, W$]	B1
	Since there is no <u>overlap</u> between the events <u>or</u> cannot happen together (o.e.)	B1
		(2)
(b)	e.g. $P(B) = \dfrac{9}{25}, \; P(T) = \dfrac{8}{25}, \; P(B \cap T) = \dfrac{5}{25}$	M1
	$P(B \cap T) \neq P(B) \times P(T)$ $[0.2 \neq 0.36 \times 0.32 = 0.1152 \text{ o.e.}]$	M1
	So B and T are <u>not</u> independent	A1cso
		(3)
(c)	$[P(W) =] \; \dfrac{7}{25}$ <u>or</u> 0.28	B1
		(1)
(d)	$[P(B \cap T) =] \; \dfrac{5}{25}$ <u>or</u> $\dfrac{1}{5}$ <u>or</u> 0.2	B1
		(1)
(e)	$[P(T \mid B) =] \; \dfrac{P(T \cap B)}{P(B)} = \dfrac{\text{"(d)"}}{(5 + 4)\big/25}$	M1
	$= \dfrac{5}{9}$ or $0.\dot{5}$	A1
		(2)
		[9]

Question Number	Scheme	Marks
5. (a)	$b = \dfrac{59.99}{33.381}$	M1
	$= 1.79713\ldots\ldots$ 1.8 or awrt 1.80	A1
	$a = 32.7 - 1.79713\ldots \times 51.83$	M1
	$= -60.44525\ldots$ awrt -60	A1
	$w = -60.445251\ldots + 1.79713\ldots l$ l and w required and awrt 2sf	A1ft
		(5)
(b)	$w = -60.445251\ldots + 1.79713\ldots \times 60$	M1
	$= 47.3825\ldots$ In range 47.3 – 47.6 inclusive	A1
		(2)
(c)	It is extrapolating so (may be) unreliable.	B1 B1
		(2)
		[9]
6. (a)	$k + 2k + 3k + 4k = 1$ or $10k = 1$	
	$k = 0.1$ (*)	B1cso
		(1)
(b)	$E(X) = 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4 = 3$	M1 A1
		(2)
(c)	$E(X^2) = 1 \times 0.1 + 4 \times 0.2 + 9 \times 0.3 + 16 \times 0.4 = 10$	M1 A1
		(2)
(d)	$\text{Var}(X) = 10 - 9 (=1)$	M1
	$\text{Var}(2 - 5X) = 5^2 \text{Var}(X) = 25$	M1 A1
		(3)
(e)	$P(1,3) + P(2,2) = 2 \times 0.1 \times 0.3 + 0.2 \times 0.2 = 0.1$ (*)	M1 A1cso
		(2)
(f)	<table><tr><td>$X_1 + X_2$</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr><tr><td>p</td><td>0.01</td><td>0.04</td><td>0.1</td><td>**0.2**</td><td>0.25</td><td>0.24</td><td>**0.16**</td></tr></table>	B1 B1
		(2)
(g)	$P(2) + P(3) = 0.05$	M1A1
		(2)
		[14]

Question Number	Scheme	Marks
7. (a)	both $\dfrac{2}{3},\dfrac{1}{3}$	B1
	$\dfrac{4}{9}$	B1
	both $\dfrac{3}{5},\dfrac{2}{5}$	B1
	all three of $\dfrac{4}{9},\dfrac{4}{9},\dfrac{5}{9}$	B1
		(4)
(b)	$P(A)=P(RR)+P(YY)=\dfrac{1}{2}\times\dfrac{2}{5}+\dfrac{1}{2}\times"\dfrac{2}{5}" \quad =\dfrac{2}{5}$	B1 M1 A1
		(3)
(c)	$P(B)=P(RRR)+P(RYR)+P(YRR)+P(YYR)$	M1
	$\left(\dfrac{1}{2}\times\dfrac{2}{5}\times"\dfrac{2}{3}"\right)+\left(\dfrac{1}{2}\times\dfrac{3}{5}\times\dfrac{5}{9}\right)+\left(\dfrac{1}{2}\times"\dfrac{3}{5}"\times\dfrac{5}{9}\right)+\left(\dfrac{1}{2}\times"\dfrac{2}{5}"\times"\dfrac{4}{9}"\right) \quad =\dfrac{5}{9}$ (*)	A1cso
		(3)
(d)	$P(A\cap B)=P(RRR)+P(YYR)$	M1
	$\left(\dfrac{1}{2}\times\dfrac{2}{5}\times\dfrac{2}{3}\right)+\left(\dfrac{1}{2}\times\dfrac{2}{5}\times\dfrac{4}{9}\right) \quad =\dfrac{2}{9}$ (*)	A1cso
		(2)
(e)	$P(A\cup B)=P(A)+P(B)-P(A\cap B)$	
	$="\dfrac{2}{5}"+\dfrac{5}{9}-\dfrac{2}{9} \quad =\dfrac{11}{15}$	
		(2)
(f)	$\dfrac{P(RRR)}{P(RRR)+P(YYY)}=\dfrac{\frac{1}{2}\times\frac{2}{5}\times\frac{2}{3}}{\left(\frac{1}{2}\times\frac{2}{5}\times\frac{2}{3}\right)+\left(\frac{1}{2}\times\frac{2}{5}\times\frac{5}{9}\right)}=\dfrac{6}{11}$	M1 A1ft A1cao
		(3)
		[17]

Examiner reports

Question 1

This proved a straightforward start to the paper. Most gave a correct tree diagram although a few oversimplified using red and not red as their outcomes and this was of no help to them in part (b). It was encouraging to see the vast majority of candidates using fractions for the probabilities and only a handful using "with replacement".

Part (b) was not answered so well with many failing to consider both cases: blue then green and green followed by blue.

Question 2

The mean was calculated accurately by the majority of the candidates but the standard deviation calculation still caused problems for many. There were few summation errors but missing square roots or failing to square the mean were some of the more common errors.

Part (b) was poorly answered. The examiners were disappointed that such a sizeable minority failed to order the list and worked quite merrily with a median larger than their upper quartile. Some worked from the total of 2757 to get quartiles of 689.5, 1378.5 and 2067.5 whilst others used cumulative totals and obtained quartiles in the thousands but still failed to see the nonsensical nature of these figures. Those who did order the list used a variety of methods to try and establish the quartiles. Whilst the examiners showed some tolerance here any acceptable method should give a median of 198 but many candidates used 186. Those who knew the rules usually scored full marks here and in parts (c) and (d).

The examiners followed through a wide range of answers in part (c) and most candidates were able to secure some marks for correctly identifying patients *B* and *F* and in part (d) for describing their skewness correctly.

Question 3

The quality of responses to this question varied considerably. Most candidates stated that $a = 0.1$ and deduced that $a + b + c = 0.7$ but many were unable to proceed sensibly beyond this point. Some assumed *a*, *b* and *c* were all equal, others had $b = 0.2$ and $c = 0.4$ and some tried to consider E(*Y*) which led to some complicated and unhelpful equations. Finding *d* produced a number of errors: some confused $P(Y = y)$ with F(*y*) and gave answers of 0.3 or 0.7 whilst others used $F(1) + F(2) + P(Y = 3)$ to obtain 0.9. Part (b) proved challenging for many. Some solved the inequality to arrive at $P(Y \geq 2)$ and others created lists of all possible values for $3Y + 2$ before deciding which cases to include. Obtaining the answer from this point should have been straightforward but some simply stopped at this point or then found $P(Y = 1) + P(Y = 2)$. Some candidates were confusing $E(3Y + 2)$ and $P(3Y + 2)$.

Question 4

In part (a) *T* and *W* or *B* and *W* were often stated as a mutually exclusive pair and a sensible reason given in words or symbols such as $P(T \cap W) = 0$. Some candidates mixed up mutually exclusive and independent events and gave *B* and *T* as their pair with a reason that they could both happen together.

The test for independence in part (b) caused problems for some. The examiners were looking for a clear application of a suitable test which was usually $P(B \cap T) \neq P(B) \times P(T)$, although some used a conditional probability successfully. We would also expect to see the required probabilities stated and this was a problem for some: a minority simply used integers and others gave $P(B) = \dfrac{4}{25}$ or $P(T) = \dfrac{3}{25}$. Those who carried out a correct test with correct probabilities usually gave a correct conclusion too and secured all 3 marks.

Part (c) was nearly always correct but a few gave $\dfrac{12}{25}$ as their answer to part (d). Most candidates now can identify a conditional probability and only a few attempted $P(B|T)$ in part (e). It was quite possible to write down this final answer from the Venn diagram but many used the conditional probability formula from the formula booklet which if their $P(B)$ was incorrect meant they could at least score the M1.

Question 5

There were some good responses to this question, but some candidates calculated the slope as 59.99/120.1, although 3/5 marks were obtained if they went on to produce the equation as $w = 6.8 + 0.50l$ provided a minimum of 2 significant figures were used. Candidates should be able to identify the independent and dependent variables from a contextual question. The accuracy mark for the calculation of the intercept was lost if they used the rounded value of 1.8 for the slope in the calculation for the intercept. Many candidates did not believe 60mm to be quite far enough away from the data range to be called extrapolation showing that they did not go back and read the question carefully enough and consider the range of values given.

Question 6

Most candidates showed us clearly that they were using the sum of the probabilities to reach $10k = 1$ from which they showed that $k = 0.1$ and the calculation of E(*X*) was usually correct too. In part (c) some confusion between E(X^2) and Var(*X*) and also over what to square (some choosing the probabilities rather than the values of *x*) caused difficulties and a few simply squared their answer to part (b). Part (d) was a fairly standard request and most knew that Var(2 – 5*X*) = 25 Var(*X*) and were sometimes able to recover from errors made in part (c) but then many solutions ground to a halt.

Many candidates did not recognise the 3 cases required in part (e) and were therefore unable to emulate this approach in part (f) but it was encouraging to see that some of these candidates did realise what was required for part (g) and often gained both marks here.

PMT

Question 7

Many candidates were able to complete the tree diagram correctly but common errors were to have probabilities of 5/9 and 4/9 on the top two and bottom two branches.

Part (b) was often answered correctly and even those with incorrect tree diagrams could achieve 2 or even all 3 marks here. Few candidates explained which four probabilities they were using to answer part (c) and the examiners were often left trying to deduce this from their tree diagram. Simply writing P(RRR) + P(RYR) + P(YRR) + P(YYR) would have earned them the first mark and made their solution much clearer. Some candidates failed to appreciate that a sum of 4 products of 3 probabilities from their tree diagram was required and a popular "fiddle" was to calculate $\frac{1}{4}\left(\frac{6}{9}+\frac{5}{9}+\frac{5}{9}+\frac{4}{9}\right)$ using the probabilities from the third branches and magically dividing by 4 to reach the printed answer. Part (d) was answered poorly with most attempts assuming that A and B were independent events (they were but this was never justified). Part (e) though was usually answered well with candidates clearly using the addition rule and the given answers.

In the final part many candidates failed to identify the conditional probability and those who did often did not clearly state what their ratio of probabilities represented: once again a statement such as $("P"(RRR))/("P"(RRR)+"P"(YYY))$ with some suitable probabilities attempted would have secured the first mark.

Statistics for S1 Practice Paper Silver Level S2

Qu	Max Score	Modal score	Mean %	ALL	A*	A	B	C	D	E	U
						Mean score for students achieving grade:					
1	5		76	3.78		4.44	3.65	3.06	2.72	2.28	1.53
2	14		71	10.00		11.92	10.10	8.94	7.93	6.86	5.07
3	7		67	4.66	6.78	6.43	5.45	4.53	3.68	3.05	2.09
4	9		64	5.75	8.44	7.91	6.66	5.62	4.68	3.92	2.93
5	9		68	6.16		7.87	6.80	6.17	5.53	4.89	3.08
6	14		60	8.43	12.50	11.13	8.56	7.25	6.24	5.12	3.28
7	17		59	10.02	14.74	13.28	10.00	7.84	6.33	5.46	3.42
	75		**65**	**48.80**		**62.98**	**51.22**	**43.41**	**37.11**	**31.58**	**21.40**