Secondary
Maths
Resources

ZigZag Education

maths@zigzageducation.co.uk
www.zigzageducation.co.uk

A LEVEL
TOPIC TESTS
STATISTICS – EDEXCEL – 2000 – S1

2nd Edition: Released 14th May 2004

Contents

1. **Data Collection and Graphical Representation**
2. **Methods for Summarising Sample Data**
3. **Probability**
4. **Correlation and Regression**
5. **Discrete Random Variables**
6. **The Normal Distribution**

ZigZag Education, Unit 3 Greenway Business Centre, Doncaster Road, Bristol, BS10 5PY.

POD1030.x-X13

Data Collection and Graphical Representation

EDEXCEL – 2000 – **S1** – syllabus reference 2

Give all your answers to 2 decimal places

1) The examination marks for a class of 20 students in 1997 are as follows:-
20, 26, 33, 38, 43, 43, 44, 48, 51, 56, 66, 67, 72, 81, 84, 90, 91, 92, 96, 97.
The examination marks for 25 students in 1998 from the equivalent class are as follows:-
12, 26, 26, 33, 33, 34, 35, 44, 51, 55, 56, 57, 61,72, 72, 73, 74, 75, 80, 81, 82, 83, 95, 95, 100
Draw a **box and whisker plot** for each set of data and **compare the distributions** and **without** doing any other **numerical** calculations comment on the **skewness** of the data. (21)
21

2) Test marks by 2 groups of students in a particular test are given below.
Group 1: 2, 12, 12, 12, 13, 14, 15,17, 19, 20, 20.
Group 2: 5, 5, 6, 6, 8, 12, 12, 15, 16, 16, 16, 16, 17, 18.
Draw a suitable **back-to-back stem and leaf diagram** for the tests results. (5)
Calculate the **median** result for the 2 groups. (2)
Without doing any other **numerical** calculations **make some comparisons** between the 2 groups. (2)
9

3) 100 web-sites were downloaded through the Internet. The frequency distribution for the time it took to download each web-site to the nearest minute was as follows:

Time	0 – 9	10 – 19	20 – 29	30 – 49
Frequency	64	22	12	2

Draw a **histogram** to represent this data and **without** doing any **numerical** calculations comment on the **skewness** of the data. (7, 1)
8

4) The amount spent on a single day by a group of 90 students was recorded and tabulated as follows:

£	0	$0 < x < 10$	$10 \le x < 20$	$20 \le x < 30$	$30 \le x < 50$	$50 \le x < 100$
Frequency	10	54	6	6	12	2

Draw a **cumulative frequency polygon** of this distribution. Estimate the **median** and the **quartiles** from your diagram.
Clearly indicate the use of the diagram to estimate these values. (12)
12

{50}

Give all your answers to 2 decimal places where appropriate

Methods for Summarising Sample Data (Basic Statistical Calculations)

EDEXCEL – **S1** – syllabus reference 2

Ordering and coding
Mean, median, mode,
Variance and sd
Interpretation
Skewness & quartiles

1) For the numbers 1, 8, 16, 24, calculate the
i) **Mean** ii) **Median** iii) **Variance** iv) **Standard Deviation** (1,1,1,1) **4**

2)

Length	10 – 19	20 – 29	30 – 49	50 – 59
Frequency	4	11	10	4

i) State the **class boundaries** and the **class widths.** (1)
ii) Calculate an **estimate** for the **mean.** (1)
iii) State the **modal class.** (1)
iv) Which class is the **median** in? (1)
v) Calculate an **estimate** for the **variance** of the data. (2) **6**

3) Data set 1 has a **mean of 2** and a **standard deviation of 4**. Data set 2 has a **mean of 3** and a **standard deviation of 4**. Data set 1 is made up of **12 numbers** and data set 2 is made up of **20 numbers**.
The 2 data sets are **combined** to make up 32 numbers.
i) Calculate the **mean** of the new combined set of numbers. (1)
ii) Calculate the **standard deviation** of the new combined set of numbers. (3)
Data set 1 has a **range** of 10 and a maximum value of 17.
Data set 2 has a **range** of 11 and a maximum value of 101.
iii) Calculate the **range** of the new **combined** set of numbers. (1) **5**

4) In Data set 1 above, the numbers are all doubled; the new set of numbers are called data set 3.
i) State the new **mean** of data set 3. (1)
ii) State the new **standard deviation** of data set 3. (1)
In Data set 2 above all the numbers have 2 added to them; the new set of numbers are called data set 4.
iii) State the new **mean** of data set 4. (1)
iv) State the new **standard deviation** of data set 4. (1) **4**

5) Here are 51 numbers summarised in table form:

Number	1	2	3	4	5	6	7	8	9
Frequency	1	2	1	3	5	6	9	10	14

i) Calculate the **quartiles** Q_1, Q_2 and Q_3. (3)
ii) Describe the **skewness** the **quartiles** indicate. (2)
Include in your answer to part ii) a numerical calculation for skewness and comment on its meaning. **5**

6) A code of the form $y = \frac{x-a}{b}$ is used to code the following information. ($a \neq 0$ and $b \neq 1$).

x	3012	3024	3036	3048	3060	3072
Frequency	3	5	15	18	1	21

Suggest **suitable** values for **a** and **b**. (2) **2**

{26}

Probability

EDEXCEL – 2000 – **S1** – syllabus reference 3

Sample space, exclusive & complimentary, conditional, independence, sum & product laws.

P(A') means the probability of not A
P(A|B) means the probability of A given that B

1) Given that A and B are 2 *independent* events with $P(A) = \frac{1}{3}$ and $P(A \text{ and } B) = \frac{1}{6}$

   i) Find P(B)
   ii) Find P(A|B)
   iii) Find P(A|B') (2,1,1)

   4

2) Given that $P(B) = \frac{1}{2}$ and $P(B \text{ and } A) = \frac{1}{10}$.
   Find P(B and A'). (2)

   2

3) Two events A and B are **mutually exclusive** and **exhaustive** and $P(A) = \frac{1}{4}$.

   i) Find P(B)
   ii) Find P(A or B)
   iii) Find P(A and B)
   iv) Find P(B|A)
   v) Find P(A|B') (1,1,1,1,1)

   5

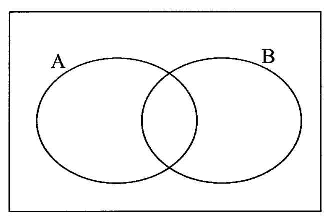
4) Two events A and B are **mutually exclusive**.
   i) Draw a **Venn diagram** to illustrate this. (2)

   2

5) Two events A and B are **exhaustive but not mutually exclusive**.
   Given $P(A) = \frac{1}{4}$ and $P(B) \neq 0$.
   State the **range** of possible values for P(A and B) (4)

   4

6) Copy this Venn diagram separately 4 times.
   i) **Shade** the area which represents P(A or B) (1)
   ii) **Shade** the area which represents P(A' or B') (1)
   iii) **Shade** the area which represents P(A' and B) (1)
   iv) **Shade** the area which represents P(A' or B) (1)
   Clearly mark your diagrams 6i) 6ii) 6iii) and 6 iv)!



   4

7) A bag contains 3 red and 4 blue balls.
   I take one ball at random and replace it back in the bag.
   I then take another ball and replace it in the bag.
   I then take another ball and replace it in the bag. In total I have taken 3 balls out of the bag and have recorded the colour of each ball.
   i) Draw a **tree diagram** to illustrate all the possible selections (2)
   ii) Calculate the probability that there are **exactly** 2 red balls selected. (2)
   iii) Calculate the probability that there are **at least** 2 red balls selected. (2)
   iv) Calculate the probability that there are **exactly** 2 red balls selected **given** that at least one ball selected is blue. (2)

   8
   {29}

Correlation
Explanatory and Response
Linear regression and
applications and interpretations

Correlation and Regression

EDEXCEL – 2000 – **S1** – syllabus reference 4

1) From experiment the following values of X and Y are recorded.
Y is time in minutes and X is the Temperature in Celsius of a rod during an experiment.

X	1	3	5	7	9	11
Y	7.5	8.5	10.4	11.6	12.5	13.9

i) Calculate the product moment correlation coefficient. (6)

ii) A student suggests that X and Y are possibly linked by the formulae $y = bx^2$

Without doing any further calculations **comment** on this suggestion and include in your own comments an improved possible formula linking X and Y. (3)

9

2) Describe the method of **least squares** for the **regression line y on** ***x*** *and include a diagram in your answer.* (3)

3

3) Two variables x and y are known to have a linear relationship. From experiment the following values are recorded.

X	1	3	5	7	9	11
Y	7.5	8.5	10.4	11.6	12.5	13.9

From the above data work out the least squares regression line y on x. (4)
Draw the graph of the linear least squares regression line y on x on suitable axes. (2)

6

4) 12 students of approximately equal ability (in recalling items) are involved in a study about sleep deprivation. The 12 students are kept awake from midnight on 21st of January. Each student is kept awake for a different amount of time. Each student is then asked to recall from memory 100 items, which they were shown the day before.
The results for 11 students are summarised below.

Hours without Sleep	0	3	6	9	12	15	18	21	24	27	30
Items Memorised	25	26	19	20	17	13	13	9	9	9	4

i) Calculate a *suitable* least squares linear regression line and state the gradient and y-intercept. (4)
ii) Give a **clear** meaning of these 2 values. (4)
The 12th student is also kept awake on 21st of January.
iii) If the student has 10 hours without sleep, how many items would you expect them to recall? Give your answer to 1 d.p. (1)
iv) Estimate the necessary hours without sleep required for the 12th student to memorise exactly 10 items. (1)
v) Explain why it is not appropriate to use the equation of the regression line to estimate the number of items recalled for 35 hours without sleep. (1)

11
{29}

Discrete RV
Probability function and CDF
Expectation

Discrete Random Variables including Expectation

EDEXCEL – 2000 – **S1** – syllabus reference 5

1) The table gives the **probability distribution** for the random variable Y:

y	0	1	2
P(Y = y)	0.01	0.495	0.495

i) Calculate E(Y) (2)
ii) Calculate Var(Y) (3)
iii) Construct the **cumulative distribution table**. (3)

8

2) The **probability distribution** for the random variable X is given by:

$$P(X = x) = ax \quad \text{for } x = 1, 2, 3, 4, 5, 6, 7, 8.$$

i) Find a. (2)
ii) Calculate E(X). (2)
iii) Calculate Var(X) (3)
iv) Calculate $P(X < 5)$ (2)
v) Calculate $P(X \leq 5)$ (2)
vi) Calculate $P(X > 5)$ (2)
vii) On any given trial, what is the most likely outcome? (2)

viii) Construct the **cumulative distribution table** and use the fact that $\sum_{1}^{n} r = \frac{1}{2}n(n+1)$ to give the cumulative distribution function (cdf) as a **formula**. (5)

Remember for completeness to include the range of x values where the cdf is zero and the range of values where the cdf is 1.

20

3) Two random variables X and Y are such that:

$$E(X) = 2 \qquad Var(X) = 3$$
$$E(Y) = 2 \qquad Var(Y) = 3$$

Calculate

i) E(4X) ii) E(3X + 4)
ii) Var (4X) iv) Var(5Y)
v) Var (5Y – 4) vi) E(2Y – 4) (12)

12
{40}

The Normal Distribution

EDEXCEL – 2000 – **S1** – syllabus reference 6

Give all your answers to 2 decimal places

1) If $Z \sim N(0, 1)$
   - i) Find $P(z < 0.6)$ (1)
   - ii) Find $P(z < -0.6)$ (1)
   - iii) Find $P(z > 0.6)$ (1)
   - iv) Find $P(0.6 < |z|)$ Draw a sketch showing this probability. (2)
   - v) Find β such that $P(z > \beta) = 0.15$ (2) **7**

2) If $X \sim N(100, 36)$
   - i) Find $P(x < 100)$ (2)
   - ii) Find $P(x < 94)$ and draw a sketch showing this probability. (2)
   - iii) Find $P(x > 91)$ and draw a sketch showing this probability. (2)
   - iv) State E(X) (1)
   - v) State Var(X) (1)
   - vi) Calculate the standard deviation. (1) **9**

3) A certain type of tomato has a mass which is **normally distributed** with mean 80g and standard deviation 25g. A truck load of 1000 such tomatoes travels down the motorway.
   - i) Estimate how many tomatoes in the truck have a mass **less than** 80g. (1)
   - ii) Estimate how many tomatoes in the truck have a mass less than 70g. (3)
   - iii) Estimate how many tomatoes in the truck have a mass **less than** 90g. (1)
   - iv) Estimate the probability that any given tomato from the truck has a **mass under** 65g. (2)

   On arrival Jane **selects 4 tomatoes** at random from the truck and checks the mass of each tomato.
   - v) Calculate the probability that **all** the tomatoes selected by Jane each weigh **less than 70g.**(2) **9**

{25}

===

The Normal Distribution

EDEXCEL – 2000 – **S1** – syllabus reference 6

Give all your answers to 2 decimal places

1) If $Z \sim N(0, 1)$
   - i) Find $P(z < 0.6)$ (1)
   - ii) Find $P(z < -0.6)$ (1)
   - iii) Find $P(z > 0.6)$ (1)
   - iv) Find $P(0.6 < |z|)$ Draw a sketch showing this probability. (2)
   - v) Find β such that $P(z > \beta) = 0.15$ (2) **7**

2) If $X \sim N(100, 36)$
   - i) Find $P(x < 100)$ (2)
   - ii) Find $P(x < 94)$ and draw a sketch showing this probability. (2)
   - iii) Find $P(x > 91)$ and draw a sketch showing this probability. (2)
   - iv) State E(X) (1)
   - v) State Var(X) (1)
   - vi) Calculate the standard deviation. (1) **9**

3) A certain type of tomato has a mass which is **normally distributed** with mean 80g and standard deviation 25g. A truck load of 1000 such tomatoes travels down the motorway.
   - i) Estimate how many tomatoes in the truck have a mass **less than** 80g. (1)
   - ii) Estimate how many tomatoes in the truck have a mass less than 70g. (3)
   - iii) Estimate how many tomatoes in the truck have a mass **less than** 90g. (1)
   - iv) Estimate the probability that any given tomato from the truck has a **mass under** 65g. (2)

   On arrival Jane **selects 4 tomatoes** at random from the truck and checks the mass of each tomato.
   - v) Calculate the probability that **all** the tomatoes selected by Jane each weigh **less than 70g.**(2) **9**

{25}

Data Collection and Graphical Representation

EDEXCEL - 2000 - **S1** - syllabus reference 2

All answers to 2 decimal places

1. 20 students in 1997, so we split the group up into two smaller groups of size 10.

Q_2 = median, use the $\frac{n+1}{2}$ formula $\Rightarrow \frac{20+1}{2} = \frac{10^{th}+11^{th}}{2} = \frac{56+66}{2} = 61$ **M1A1**

Q_1 is the median of the first group $\Rightarrow \frac{10+1}{2} = \frac{5^{th}+6^{th}}{2} = \frac{43+43}{2} = 43$ **M1 A1 #**

Q_3 is the median of the second group $\Rightarrow \frac{15^{th}+16^{th}}{2} = \frac{84+90}{2} = 87$ **M1 A1 #**

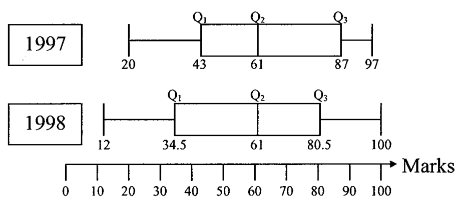
25 students in 1998; this has a middle number, the 13th mark, which is Q_2. We then ignore that one and split the rest of the data into two smaller groups of size 12. $Q_2 = 61$ **M1 A1**

Q_1 is the median of the first group $\Rightarrow \frac{6^{th}+7^{th}}{2} = \frac{34+35}{2} = 34.5$ **M1 A1 #**

Q_3 is the median of the second group $\Rightarrow \frac{19^{th}+20^{th}}{2} = \frac{80+81}{2} = 80.5$ **M1 A1 #**

Note: Using the method in the new Heinemann S1 text, 1998 quartiles are Q1 = 35, Q3 = 81

> Draw two box and whisker plots, one each for 1997 Class and 1998 Class.
> 1997: left tail at 20, Q_1, Q_2, Q_3 as above, right tail at 97 **A1**
> 1998: left tail at 12, Q_1, Q_2, Q_3 as above, right tail at 100 **A1**
> Draw common scale for both, in increments of 10, labelled "Marks"
> Mark scheme: Correct range (× 2) **M1**, Scale **M1**, Common Scale **M1**



The range of the 1998 data is larger **A1**, although the main bulk of the marks has a similar **A1** spread over both years. This you can tell from the similar widths of Q_3 - Q_1.
The 1997 set is better and has positive **A1** skewness.
The 1998 set has negative **A1** skewness.

or some other *correct* method for calculating Q_1 and Q_3 (many methods exist!) **21**

2.

Group 1		Group 2		
1	1		appropriate stem	**M2**
	5	0 0 1 1 3	correct leafs	**A1**
3 3 3	9	3 3	back to back	**M1**
2 1 0	13	2 3 3 3 3		
3 3 2 0	17	0 1		

Key: 1 | 1 means 2 **M1**

Median of Group 1 = 14 **A1** Median of Group 2 = 13.5 **A1**

Any 2 correct statements comparing the two distributions **A1 A1**

e.g. Group 1 is stronger as it has a higher median.
Group 1 is uniformly distributed at the top end.
Group 2 has weaker students as is shown by the peak under stem 5. **9**

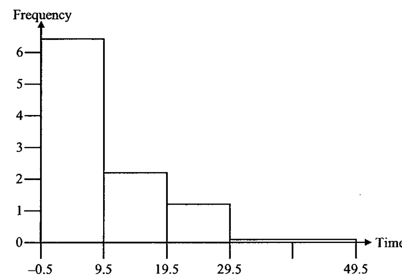
3.

Time	0 - 9	10 - 19	20 - 29	30 - 49
Frequency	64	22	12	2
Column height	6.4 *(6.74)	2.2	1.2	0.1
Column width	10 *(9.5)	10	10	20

M1 for column height, **M1** for column width
* Also correct is treating first column width as 9.5, i.e. (0 - 9.5)

Histogram to represent data above, with appropriate title.
M1 for appropriate scale on frequency axis
M1 for correct scale on time axis
B1 for label on frequency axis
B1 for correct height of first three columns
B1 for correct height of fourth column

A1 for stating positive skewness



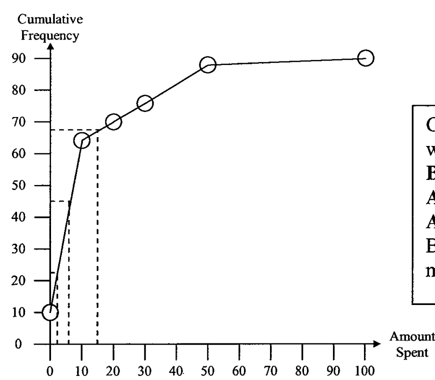
8

4.

Upper class boundary	0	10	20	30	50	100
Cumulative frequency	10	64	70	76	88	90

↑**B1**

M1 for each row (or implied) i.e. M1 M1 for both rows



Cumulative frequency diagram of above data, with appropriate title.
B1 for label on cumulative frequency axis
A1 for first point (0, 10)
A1 for other correct points
Broken lines on graph showing estimation of median and quartiles

Q_1 **estimated at C.F. = 22.5 ⇒ 2** **M1 A1**
Q_2 estimated at C.F. = 45 ⇒ 6 **M1 A1**
Q_3 estimated at C.F. = 67.5 ⇒ 16 **M1 A1**

12

{50}

Methods for Summarising Sample Data
(Basic Statistical Calculations)

EDEXCEL - 2000 - **S1**

All answers to 2 decimal places where appropriate

1. i) Mean = $\frac{1+8+16+24}{4} = 12.25$ **A1** ii) Median = $\frac{8+16}{2} = 12$ **A1**

   iii) Variance = $\frac{\sum x^2}{n} - \bar{x}^2 = \frac{897}{4} - (12.25)^2 = 74.1875 = 74.19$ (2 d.p.) **A1**

   iv) S.D. = $\sqrt{Variance} = \sqrt{74.1875} = 8.61$ **A1** **4**

2. i) Class boundaries: 9.5, 19.5, 29.5, 49.5, 59.5
   Class widths: 10, 10, 20, 10 respectively **A1**

   ii) Mean $\approx \frac{(4\times 14.5)+(11\times 24.5)+(10\times 39.5)+(4\times 54.5)}{29} = 32.43$ **A1**

   iii) 20 – 29 **A1** iv) $\frac{29}{2} = 14.5 \Rightarrow 20 - 29$ **A1**

   v) $\sum x^2 = 4(14.5)^2 + 11(24.5)^2 + 10(39.5)^2 + 4(54.5)^2$ **M1** for using mid-points
   $= 34927.25$

   Variance = $\frac{\sum x^2}{n} - \bar{x}^2 = \frac{34927.25}{29} - \left(\frac{940.5}{29}\right)^2 = 152.62$ **A1** **6**

3. i) $\bar{x} = \frac{\sum x}{n} \Rightarrow 2 = \frac{\sum x}{12} \Rightarrow \sum x_1 = 24.$ $\bar{x} = \frac{\sum x}{n} \Rightarrow 3 = \frac{\sum x}{20} \Rightarrow \sum x_2 = 60$

   $\therefore \bar{x} = \frac{\sum x_1 + \sum x_2}{n_1 + n_2} = \frac{24+60}{32} = 2.625$ or 2.63 (2 d.p.) **A1**

   ii) Variance of Data set 1= $\frac{\sum x^2}{n} - \bar{x}^2 \Rightarrow 16 = \frac{\sum x^2}{12} - 4 \Rightarrow \sum x^2 = 240$ **B1**

   Variance of Data set 2 = $\frac{\sum x^2}{n} - \bar{x}^2 \Rightarrow 16 = \frac{\sum x^2}{20} - 9 \Rightarrow \sum x^2 = 500$ **B1**

   Variance of combined set = $\frac{240+500}{32} - 2.625^2 = 16.234375$

   S.D. = $\sqrt{Variance} = 4.03$ **A1**

   iii) Set 1, range 10, max 17 $\Rightarrow$ min 7.
   Set 2, range 11, max 101 $\Rightarrow$ min 90.
   $\therefore$ Range = 101 – 7 = 94 **A1** **5**

4. i) $2 \times 2 = 4$ **A1** ii) $2 \times 4 = 8$ **A1**
   iii) $3 + 2 = 5$ **A1** iv) 4 **A1** **4**

5. i) $Q_1 = 13^{th}$ number = 6 **A1** # ii) $\frac{Q_1 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{9-2(7)+6}{9-6} = \frac{1}{3}$

   $Q_2 = 26^{th}$ number = 7 **A1** **A1** or equivalent
   $Q_3 = 39^{th}$ number = 9 **A1** # $\therefore$ positively skewed **A1**
   # or other correct quartiles **5**

6. a = answer from 3000, 3012, 3024, 3036, 3048, 3060, 3072. **A1**
   b = 12 **A1** **2**

{24}

Probability

EDEXCEL - 2000 - **S1** - syllabus reference 3

P(A') means the probability of not A

P(A|B) means the probability of A given B

1. i) $P(B) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$ **M1 A1**

   ii) $P(A|B) = P(A) = \frac{1}{3}$ **A1**

   iii) $P(A|B') = P(A) = \frac{1}{3}$ **A1** **4**

2. $P(B) = P(B \cap A) + P(B \cap A')$ **M1**

   $\therefore P(B \cap A') = P(B) - P(B \cap A) = \frac{1}{2} - \frac{1}{10} = \frac{2}{5}$ **A1** **2**

3. i) $P(B) = \frac{3}{4}$ **A1**

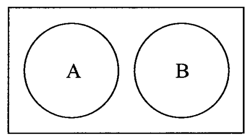
   ii) P(A or B) = 1 **A1**

   iii) P(A and B) = 0 **A1**

   iv) $P(B|A) = 0$ **A1**

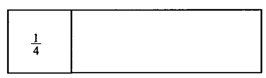
   v) $P(A|B') = 1$ **A1** **5**

4. Venn diagram to illustrate two mutually exclusive events: Two non-intersecting circles in a rectangle, circles labelled A and B **A2**



2

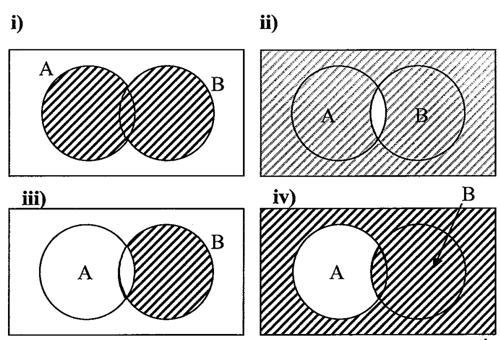
5. $0 <$ **A2** $P(A \cap B) \leq$ **A2** $\frac{1}{4}$



4

6. i), ii), iii), iv)

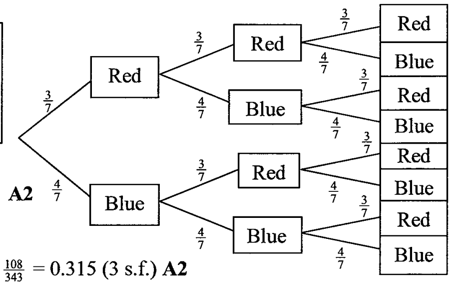
Four Venn diagrams here, each with the same basic layout: two intersecting circles, labelled A and B, inside a rectangle.
i) P(A or B): Interior of both circles shaded **A1**
ii) P(A' or B'): Whole diagram shaded *except* for intersection of two circles **A1**
iii) P(A' and B): Circle B shaded *except* for intersection of two circles **A1**
iv) P(A' or B): Whole diagram shaded *except* for the part of circle A that does not intersect circle B **A1**
Clearly mark diagrams 6 i), 6 ii), 6 iii) and 6 iv)!



4

7. i)

Tree diagram, branching 3 times to give 8 final branches, each branching point having a probability of $\frac{3}{7}$ of selecting a red ball and $\frac{4}{7}$ of selecting a blue ball.



A2

ii) P(exactly 2 red balls selected) = $3\left(\frac{3}{7}\right)\left(\frac{3}{7}\right)\left(\frac{4}{7}\right) = \frac{108}{343} = 0.315$ (3 s.f.) **A2**

iii) P(at least 2 red balls selected) = $\frac{108}{343} + \left(\frac{3}{7}\right)^3 = \frac{135}{343} = 0.394$ (3 s.f.) **M1 A1**

iv) P(2 red | ≥ 1 blue) = $\dfrac{P(2\text{ red and }1\text{ blue})}{P(\geq 1\text{ blue})} = \dfrac{\frac{108}{343}}{1-\frac{27}{343}} = \frac{108}{316} = 0.342$ (3 s.f.) **M1 A1**

8

{29}

Correlation and Regression

EDEXCEL - 2000 - **S1** - syllabus reference 4

1.

X	1	3	5	7	9	11	**36**
Y	7.5	8.5	10.4	11.6	12.5	13.9	**64.4**
XY	7.5	25.5	52	81.2	112.5	152.9	**431.6**
X^2	1	9	25	49	81	121	**286**
Y^2	56.25	72.25	108.16	134.56	156.25	193.21	**720.68**

i) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{36}{6} = 6,\ \bar{y} = \dfrac{\sum y}{n} = \dfrac{64.4}{6} = \dfrac{161}{15} = 10.73$ (2 d.p.)

$S_{xy} = \dfrac{\sum xy}{n} - \bar{x}\bar{y} = \dfrac{431.6}{6} - 6\left(\dfrac{161}{15}\right) = \dfrac{1079}{15} - 64.4 = \dfrac{113}{15} = 7.53$ (2 d.p.)

$S_x^2 = \dfrac{\sum x^2}{n} - \bar{x}^2 = \dfrac{286}{6} - 6^2 = \dfrac{35}{3} = 11.67$ (2 d.p.)

$S_y^2 = \dfrac{\sum y^2}{n} - \bar{y}^2 = \dfrac{720.68}{6} - \left(\dfrac{161}{15}\right)^2 = \dfrac{18017}{150} - \dfrac{25921}{225} = \dfrac{2209}{450} = 4.91(2\ d.p.)$

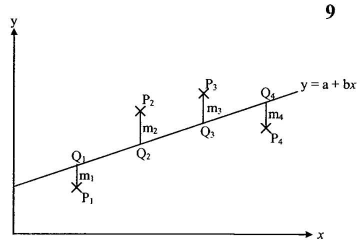
$r = \dfrac{S_{xy}}{S_x S_y} = \dfrac{\frac{113}{15}}{\sqrt{\frac{35}{3}}\sqrt{\frac{2209}{450}}} = 0.995456275 = 0.9955$ (4 d.p.) **M3 A3**

ii) The data are clearly linked in a linear manner (i.e. r close to 1)

$\therefore$ $y \neq bx^2$ **A1** and $\therefore$ suggest $y = ax$ **A1**+ b **A1** **9**

2.

Draw scatter diagram: axes labelled x and y, line y = a + bx (a > 0, 0 < b < 1), four scatter points (little x's) labelled P_i, i = 1, 2, . . . n, (n = 4 here), two below line, two above line, vertical lines from scatter points to line, lengths m_1, m_2, m_3, m_4, points on line where these lines meet y = a + bx, labelled Q_i **A1**



Drawn on the scatter diagram, this shows the points P_i (x_i, y_i) where i = 1, 2, . . ., n.
The lengths P_1Q_1, P_2Q_2, . . . , P_nQ_n are residuals and denoted by m_1, m_2, . . . , m_n.
$\sum m_i^2$ is the sum of the squares of the residuals. It is possible to find values of a and b such that $\sum m_i^2$ **A1** is a minimum **A1**. With these values, the line y = a + bx is known as the least squares regression line on x. **3**

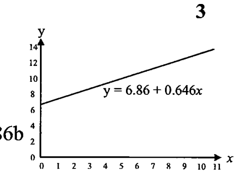
3. $\sum x = 36 \quad \sum y = 64.4 \quad \sum x^2 = 286 \quad \sum xy = 431.6\ n = 6$

$\sum y = na + b\sum x \Rightarrow 64.4 = 6a + 36b \Rightarrow -386.4 = -36a - 216b$

$\sum xy = a\sum x + b\sum x^2 \qquad \Rightarrow 431.6 = 36a + 286b$

$\therefore 45.2 = 70b \therefore b = \frac{113}{175} = 0.64571$ (5 d.p.) $\Rightarrow a = 6.85905$ (5 d.p.)

$\therefore$ y = 6.86 + 0.646x (3 s.f.) **M2 A2**



Draw graph, axes labelled x (0 - 11, step length 1) and y (0 - 14, step length 2). Graph of y = 6.86 + 0.646x **A2**

6

4. i)

x	Hours without sleep	0	3	6	9	12	15	18	21	24	27	30	**165**
y	Items memorised	25	26	19	20	17	13	13	9	9	9	4	**164**
x^2		0	9	36	81	144	225	324	441	576	729	900	**3465**
y^2		625	676	361	400	289	168	169	81	81	81	16	**2948**
xy		0	78	114	180	204	195	234	189	216	243	120	**1773**

$\sum y = \text{na} + \text{b}\sum x \Rightarrow 164 = 11\text{a} + 165\text{b} \Rightarrow -2460 = -165\text{a} - 2475\text{b}$

$\sum xy = \text{a}\sum x + \text{b}\sum x^2 \qquad \Rightarrow 1773 = 165\text{a} + 3465\text{b}$

$\therefore -687 = 990\text{b} \;\therefore\; \text{b} = -\frac{229}{330} = -0.69\dot{3}\dot{9} \qquad \therefore\; \text{a} = \frac{557}{22} = 25.3\dot{1}\dot{8}$

$y = 25.3\dot{1}\dot{8} - 0.69\dot{3}\dot{9}x$ **M2 A2**

Gradient $= -0.69\dot{3}\dot{9}$

Y-intercept $= 25.3\dot{1}\dot{8}$

ii) Y-intercept: - Expected number of items recalled by student without sleep deprivation.**A2**
Gradient: - For every extra hour of sleep deprivation, the reduction in the number of items a student is expected to recall. **A2**

iii) $y = 25.3\dot{1}\dot{8} - 0.69\dot{3}\dot{9}(10) = 18.4$ (1 d.p.) **A1**

iv) $10 = 25.3\dot{1}\dot{8} - 0.69\dot{3}\dot{9}x \Rightarrow x = \dfrac{10 - 25.3\dot{1}\dot{8}}{-0.69\dot{3}\dot{9}} = 22.07$ hrs (2 d.p.) **A1**

v) Extrapolation is dangerous outside original range of hours without sleep. **A1** **11**

{20}

Discrete Random Variables

EDEXCEL - 2000 - **S1** - syllabus reference 5

1. i) $E(Y) = \sum_{all\,y} yP(Y=y) = 0(0.01) + 1(0.495) + 2(0.495) = 1.485$ **M1 A1**

   ii) $Var(Y) = \sum y_i^2 p_i - \bar{y}^2 = (0(0.01) + 1(0.495) + 4$ **B1**$(0.495)) - (1.485)^2$ **M1**

   $= 0.269775 = 0.2698$ (4 d.p.) **A1**

   iii)

y	0	1	2
$P(Y \le y)$	0.01	0.505	1
	A1	**A1**	**A1**

8

2. i) $\sum_{all\,x} P(X=x)$ **M1** $= a + 2a + 3a + 4a + 5a + 6a + 7a + 8a \Rightarrow 36a = 1$

   $\therefore a = \frac{1}{36}$ **A1**

   ii) $E(X) = \sum_{all\,x} xP(X=x) = \frac{1}{36}(1 + 4 + 9 + 16 + 25 + 36 + 49 + 64) = \frac{204}{36} = \frac{17}{3}$ **M1 A1**

   iii) $Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{36}(1 + 8 + 27 + 64 + 125 + 216 + 343 + 512) - (\frac{17}{3})^2$ **M1**

   $= 36 - \frac{289}{9} = \frac{35}{9}$ **A2**

   iv) $P(X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36}$ **M1**

   $= \frac{10}{36}$ $(= \frac{5}{18})$ **A1**

   v) $P(X \le 5) = \frac{10}{18} + P(X = 5)$ **M1** $= \frac{10}{36} + \frac{5}{36} = \frac{15}{36}$ $(= \frac{5}{12})$ **A1**

   vi) $P(X > 5) = 1 - P(X \le 5)$ **M1** $= \frac{21}{36}$ $(= \frac{7}{12})$ **A1**

   vii) 8 **A2**

   viii)

x	1	2	3	4	5	6	7	8
$P(X \le x)$	$\frac{1}{36}$	$\frac{3}{36}$ $(= \frac{1}{12})$	$\frac{6}{36}$ $(= \frac{1}{6})$	$\frac{10}{36}$ $(= \frac{5}{18})$	$\frac{15}{36}$ $(= \frac{5}{12})$	$\frac{21}{36}$ $(= \frac{7}{12})$	$\frac{28}{36}$ $(= \frac{7}{9})$	$\frac{36}{36}$ $(= 1)$

A2

$$P(X \le x) = \begin{cases} 0 & x < 1 \quad \textbf{A1} \\ \frac{x(x+1)}{72} & x = 1, 2, 3, 4, 5, 6, 7, 8 \quad \textbf{A1} \\ 1 & \text{otherwise} \quad \textbf{A1} \end{cases}$$

20

3. i) $E(4X) = 4E(X) = 8$ **A2**

   ii) $E(3X + 4) = 3E(X) + 4 = 10$ **A2**

   iii) $Var(4X) = (4)^2 Var(X) = 48$ **A2**

   iv) $Var(5Y) = (5)^2 Var(Y) = 75$ **A2**

   v) $Var(5Y - 4) = 75$ **A2**

   vi) $E(2Y - 4) = 2E(Y) - 4 = 0$ **A2**

12

{40}

The Normal Distribution

EDEXCEL - 2000 - **S1** - syllabus reference 6

All answers to 2 decimal places

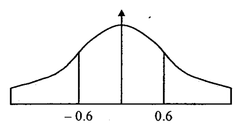
1. i) $P(Z < 0.6) = 0.7257 = 0.73$ (2 d.p.) **A1**

ii) $P(Z < -0.6) = 1 - P(Z < 0.6) = 1 - 0.7257 = 0.2743 = 0.27$ (2 d.p.) **A1**

iii) $P(Z > 0.6) = 0.2743 = 0.27$ (2 d.p.) **A1**

iv) $P(0.6 < |Z|) = 2(0.2743) = 0.5486 = 0.55$ **A1**

Sketch of Normal Distn.: Z ~ N(0,1), centred on y axis, vertical lines drawn at ± 0.6, labelled as such, area under tails of graph shaded. **A1**



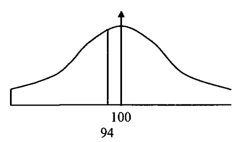
v) $P(Z > \beta) = 0.15 \Rightarrow P(Z < \beta) = 1 - 0.15 = 0.85 \quad \therefore \beta = 1.0364 = 1.04$ (2 d.p.) **A2** **7**

2. i) $P(X < 100) = P\left(\frac{X-100}{6} < 0\right) = P(Z < 0) = 0.5$ **A2**

ii) $P(X < 94) = P\left(\frac{X-100}{6} < \frac{94-100}{6}\right)$

$= P(Z < -1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587 = 0.16$ (2 d.p.) **A1**

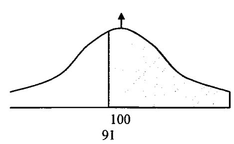
Sketch of Normal Distn.: X ~ N(100, 36), centred on $x = 100$ vertical line drawn at 94, labelled as such, area under graph left of 94 shaded. **A1**



iii) $P(X > 91) = P\left(\frac{X-100}{6} > \frac{91-100}{6}\right)$

$= P(Z > -1.5) = \Phi(1.5) = 0.9332 = 0.93$ (2 d.p.) **A1**

Sketch of Normal Distn.: X ~ N(100, 36), centred on $x = 100$ vertical line drawn at 91, labelled as such, area under graph right of 91 shaded. **A1**



iv) $E(X) = 100$ **A1**

v) $Var(X) = 36$ **A1**

vi) S.D. $= \sqrt{36} = 6$ **A1** **9**

3. i) 500 **A1**

ii) $P(X < 70) = P\left(\frac{X-80}{25} < \frac{70-80}{25}\right)$ **M1**

$= P(Z < -0.4) = 1 - \Phi(0.4) = 1 - 0.6554 = 0.3446$

$\therefore$ No. of tomatoes $= 0.3446 \times 1000$ **M1** $= 344.6$ **A1**

iii) $P(X < 90) = P\left(\frac{X-80}{25} < \frac{90-80}{25}\right) = P(Z < 0.4) = 0.6554$

$\therefore$ No. of tomatoes $= 0.6554 \times 1000 = 655.4$ **A1**

(OR $1000 - 344.6 = 655.4$)

iv) $P(X < 65) = P\left(\frac{X-80}{25} < \frac{65-80}{25}\right)$

$= P(Z < -0.6) = 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743 = 0.27$ (2 d.p.) **M1 A1**

v) $P(X < 70) = 0.3446$ $\therefore$ For 4 tomatoes: $(0.3446)^4$ **M1** $= 0.014101363 = 0.01$ **A1** **9**

{25}