|  |  |
| --- | --- |
| **Pearson Edexcel Level 3** | |
| **GCE Mathematics**  **Advanced Subsidiary**  **Paper 2A: Statistics** | |
| **Specimen paper**  **Time: 35 minutes** | **Paper Reference(s)** |
| **8MA0/02A** |
| **You must have:**  **Mathematical Formulae and Statistical Tables, calculator** | |

**Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for algebraic manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.**

**Instructions**

• Use black ink or ball-point pen.

• If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).

• Fill in the boxes at the top of this page with your name, centre number and candidate number.

• Answer **all** the questions in Section A.

• Answer the questions in the spaces provided – *there may be more space than you need*.

• You should show sufficient working to make your methods clear. Answers without working may not gain full credit.

• Inexact answers should be given to three significant figures unless otherwise stated.

**Information**

• A booklet ‘Mathematical Formulae and Statistical Tables’ is provided.

• There are 4 questions in this section. The total mark for Part A of this paper is 30.

• The marks for each question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

**Advice**

• Read each question carefully before you start to answer it.

• Try to answer every question.

• Check your answers if you have time at the end.

• If you change your mind about an answer, cross it out and put your new answer and any working underneath.

**SECTION A: STATISTICS**

**Answer ALL questions.**

**1.** A company manager is investigating the time taken, *t* minutes, to complete an aptitude test. The human resources manager produced the table below of coded times, *x* minutes, for a random sample of 30 applicants.

|  |  |  |
| --- | --- | --- |
| **Coded time**  **(*x* minutes)** | **Frequency (f )** | **Coded time midpoint**  **(*y* minutes)** |
| 0 ≤ *x* < 5 | 3 | 2.5 |
| 5 ≤ *x* < 10 | 15 | 7.5 |
| 10 ≤ *x* < 15 | 2 | 12.5 |
| 15 ≤ *x* < 25 | 9 | 20 |
| 25 ≤ *x* < 35 | 1 | 30 |

(You may use ∑*fy* = 355 and ∑*f y* 2 = 5675)

(a) Use linear interpolation to estimate the median of the coded times.

**(2)**

(b) Estimate the standard deviation of the coded times.

**(2)**

The company manager is told by the human resources manager that he subtracted 15 from each of the times and then divided by 2, to calculate the coded times.

(c) Calculate an estimate for the median and the standard deviation of *t*.

**(3)**

The following year, the company has 25 positions available. The company manager decides not to offer a position to any applicant who takes 35 minutes or more to complete the aptitude test.

The company has 60 applicants.

(d) Comment on whether or not the company manager’s decision will result in the company being able to fill the 25 positions available from these 60 applicants. Give a reason for your answer.

**(2)**

**(Total for Question 1 is 9 marks)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2.** The discrete random variable *X* ~ B(30, 0.28)

(a) Find P(5 ≤ *X* < 12).

**(2)**

Over a long period of time it has been found that 5% of the articles made by a particular manufacturer have been defective.

A company taking delivery of a large batch of these articles uses the following procedure in order to decide whether to accept the batch. The batch is accepted if:

• A random sample of 10 articles from the batch contains fewer than 2 defective articles

OR

• If the sample of 10 articles contains exactly two defective articles, a second random sample of 10 is found to contain no defective articles.

(b) What is the probability that the company will reject the delivered batch?

**(6)**

**(Total for Question 2 is 8 marks)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**3.** Pete is investigating the relationship between daily rainfall, *w* mm, and daily mean pressure, *p* hPa, in Perth during 2015. He used the large data set to take a sample of size 12.

He obtained the following results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *p* | 1007 | 1012 | 1013 | 1009 | 1019 | 1010 | 1010 | 1010 | 1013 | 1011 | 1014 | 1022 |
| *w* | 102.0 | 63.0 | 63.0 | 38.4 | 38.0 | 35.0 | 34.2 | 32.0 | 30.4 | 28.0 | 28.0 | 15 |

Pete drew the following scatter diagram for the values of *w* and *p* and calculated the quartiles.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Q1 | Q2 | Q3 |
| *p* | 1010 | 1011.5 | 1013.5 |
| *w* | 29.2 | 34.6 | 50.7 |

An outlier is a value which is more than 1.5 times the interquartile range above Q3 or more than 1.5 times the interquartile range below Q1.

(a) Show that the 3 points circled on the scatter diagram above are outliers.

**(2)**

(b) Describe the effect of removing the 3 outliers on the correlation between daily rainfall and daily mean pressure in this sample.

**(1)**

John has also been studying the large data set and believes that the sample Pete has taken is not random.

(c) From your knowledge of the large data set, explain why Pete’s sample is unlikely to be a random sample.

**(1)**

John finds that the equation of the regression line of w on p, using all the data in the large data set, is

*w* = 1023 – 0.223*p*

(d) Give an interpretation of the figure –0.223 in this regression line.

**(1)**

John decided to use the regression line to estimate the daily rainfall for a day in December when the daily mean pressure is 1011 hPa.

(e) Using your knowledge of the large data set, comment on the reliability of John’s estimate.

**(1)**

**(Total for Question 3 is 6 marks)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**4.** Alyona, Dawn and Sergei are sometimes late for school.

The events *A*, *D* and *S* are as follows:

*A* Alyona is late for school

*D* Dawn is late for school

*S* Sergei is late for school

The Venn diagram below shows the three events *A*, *D* and *S* and the probabilities associated with each region of *D*. The constants *p*, *q* and *r* each represent probabilities associated with the three separate regions outside *D*.

0.1

0.05

*r*

*A*

*D*

*S*

(a) Write down 2 of the events *A*, *D* and *S* that are mutually exclusive. Give a reason for your answer.

**(1)**

The probability that Sergei is late for school is 0.2. The events *A* and *D* are independent.

(b) Find the value of *r*.

**(4)**

Dawn and Sergei’s teacher believes that when Sergei is late for school, Dawn tends to be late for school.

(c) State whether or not *D* and *S* are independent, giving a reason for your answer.

**(1)**

(d) Comment on the teacher’s belief in the light of your answer to part (c).

**(1)**

**(Total for Question 4 is 7 marks)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**TOTAL FOR THIS PAPER IS 30 MARKS**