Candidate surname		v before ente	Other names
Pearson Edexcel nternational Advanced Level	Centr	e Number	Candidate Number
Wednesday 1	13 J	anu	ary 2021
Afternoon (Time: 1 hour 30 minu	utes)	Paper Re	eference <b>WST01/01</b>
Mathematics International Advance	ed Sul	bsidiar	y/Advanced Level
Statistics S1			

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic algebra 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 questions and ensure that your answers to parts of questions are clearly labelled.
- 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.
- Values from the statistical tables should be quoted in full. If a calculator is used instead of the tables, the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

#### Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 75.
- 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.

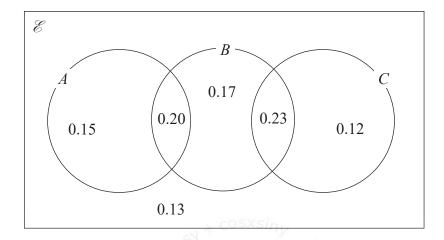
  Turn over





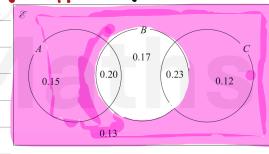
#### probability with Venn diagrans

1. The Venn diagram shows the events A, B and C and their associated probabilities.



Find

- (a) **P**(B')
- (1)
- (b)  $P(A \cup C)$
- condition (2)
- (c) P(A|B') = -b + 16 4ac
- (a) the set notation (B') means consider everything not in B
  i.e everything NOT inside the entire full circle of A
  4 because the Venn diagram only has probabilities, we can
  read Straight off the given Venn diagram



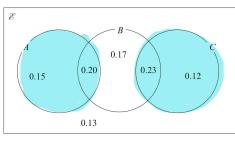
$$\frac{1}{12} \rho(8) = 0.15 + 0.13 + 0.12$$

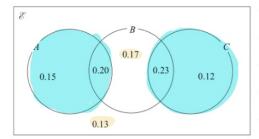
- (b) the set notation P(AVC) means A or C i.e everything that is EITHER in A OR C
- WAY 1: we shade in A and then C WAY 2: Use 1-(AVC')



Question 1 continued

### and merge





$$P(AVC) = 1 - (0.13 + 0.17)$$
  
= 0.7

(c) we see that we have a conditional probability

4 p(A 18) - there are 2 ways to tackle this:

WAY 1: visually

we need to find P(Anb'16')

i.e first narrow the Venn

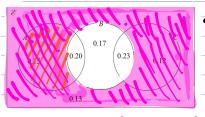
diagram to what you're

conditioning on - P(b') from
(a)

8
0.15
0.20
0.17
0.23
0.12

| becomes the denom. of the prob.

then consider out of the things in B', how many are in ANB'i.e everything that is in A
AND AT THE SAME TIME in B'
.. the overlapping region



· P(Ane'16') = 0.15/0.4 = 3/8

0.15

becomes the numer. of the prob. WAY 2: using the conditional probability formula:

formula:  $P(AIB') = P(A \cap B')$  part(a)

...first, start with the denom:

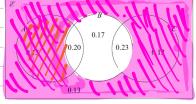
we know that P(B') is 0.4, from part (a)

... next, look at the numerator:
ue need to look for the intersection
between A and B'

i.e everything that is in A

AND AT THE SAME TIME in B'

.. the overlapping region



O.15

(becomes
the numer.
of the prob.

·· P(Ang' 6') = 0.15/0.4 = 3/8

(Total 5 marks)

Q1

blank

Leave

#### Data Representations - stem and leaf,

#### measures of central tendency, outliers,

#### box plots

2. The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

A	ge							K	ey:	4   3 i	is an a	ige of 43
4	3											(1)
5	4											(1)
6	2	3	5	6	8	8	8	9	9			(9)
7	1	1	3	4	4	6	6	6		8	9	(11)
8	0	0	2	7	8	8	9					(7)
9	3	7										(2)

(a) Find the median age of the residents.

**(1)** 

(b) Find the interquartile range (IQR) of the ages of the residents.

**(2)** 

An outlier is defined as a value that is either

more than  $1.5 \times (IQR)$  below the lower quartile or

more than  $1.5 \times (IQR)$  above the upper quartile.

- (c) Determine any outliers in these data. Show clearly any calculations that you use.
  - (3)
- (d) On the grid on page 5, draw a box plot to summarise these data.

**(3)** 

(a) finding the median age means finding the middle age value

4 summing up all the ages: 1+1+9+11+7+2 = 29

and subbing into the formula for median:

formula: 
$$\frac{n}{2} = \frac{29}{2} = 14.5 \text{ th value}$$

hremember that if you have an odd number of values, you can round up

"median = 15th value

4 reading off the stem and leaf diagram

(b) now we're asked for the IQR=Q2-Q1

sub into the formula for Q:



#### Question 2 continued

4 remember that if you have an odd number of values, you can round up

hreading off the stem and leaf diagram

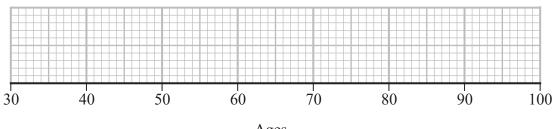
sub into the formula for Q2:

hremember that if you have an odd number of values, you can round up

hreading off the stem and leaf diagram

subbing into the IRR formula:

(c) subbing into the given formulae for outliers:



Ages

(Total 9 marks)

Q2

68-1.5(12) = 50

=) 43 is an outlier as 43250

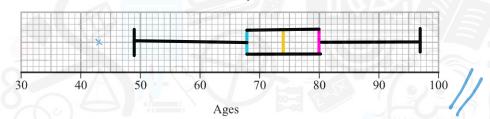
$$(a_3+1.5(12)) = 98$$

=> no outliers as all ages 498

: 43 is the only outlier

# (d) remembering the features of a box plot:

- ·outliers 43
- · lower value 54
- · lower quartile-68
- · median 74
- · upper quartile-80
- · greater values-97
- · outliers-n/a



My Maths Cloud

Normal Distribution - finding probabilities,

#### inverse normal, independent events

3. The weights of packages that arrive at a factory are normally distributed with a mean of 18 kg and a standard deviation of 5.4 kg

(a) Find the probability that a randomly selected package weighs less than 10 kg

**(3)** 

Leave

blank

The heaviest 15% of packages are moved around the factory by Jemima using a forklift truck.

(b) Find the weight, in kg, of the lightest of these packages that Jemima will move.

**(3)** 

One of the packages **not** moved by Jemima is selected at random.

(c) Find the probability that it weighs more than 18kg

**(4)** 

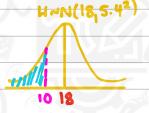
A delivery of 4 packages is made to the factory. The weights of the packages are independent.

(d) Find the probability that exactly 2 of them will be moved by Jemima.

(3)

# (a) W~N(18, 5.42)

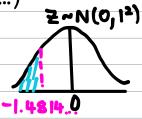
and we're asked for P(H410)



# standardise to use the prob. tables:

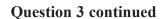
sub into the above

= P(ZL-1.481481...)

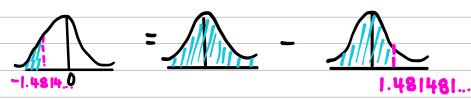


but we know that the probability tables in the formula booklet only give us probabilities for P(242) and for the tre tail

4 hence, we need to find:







NOTE: on CLASSUIZ calc, you can use the NORMAL C.D, then have 10 as the upper value to get 0.0694

(b) the question is asking for the u such that

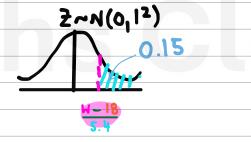
$$P(V > W) = 0.15, i.e.$$

4 inverse normal

H~N(18,5.42)

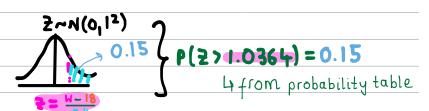


Standardize in order to use the probability table



this looks like a simple, integer probability-hence a sign to use the smaller % point table, where prob. are given for P(272)

thence we need to use:





```
Question 3 continued
```

×5.4

x5.4

w-18=5.59656

=) 
$$W = 23.59656$$
  
= 23.6 kg (35.f)

NOTE: on classure, use Inverse NORMAL with p= 0.85 (always cumulative prob.)

(c) it's important to notice that we're asked to find a conditional probability, with the condition being that w 423.59656.

(heavier packages will be picked up by Jemima 4 hence, we're asked to find:

4 subbing this into the conditional prob. formula:

.- first, the denominator:

... next, the numerator:



Question 3 continued

this makes = 
$$\frac{0.35}{0.85} = \frac{35^{7}}{85^{17}}$$

these are multiplied as are independent

4 and there are 4C2 combinations of the above:

$$(0.15)^2(0.85)^2 \times 6$$

Q3

(Total 13 marks)



**4.** A spinner can land on the numbers 10, 12, 14 and 16 only and the probability of the spinner landing on each number is the same. The random variable *X* represents the number that the spinner lands on when

it is spun once.

(a) State the name of the probability distribution of X.

(1)

(b) (i) Write down the value of E(X)

**(1)** 

(ii) Find Var(X)

**(2)** 

A second spinner can land on the numbers 1, 2, 3, 4 and 5 only. The random variable *Y* represents the number that this spinner lands on when it is spun once. The probability distribution of *Y* is given in the table below

у	1	2	3	4	5
P(Y=y)	$\frac{4}{30}$	9 30	$\frac{6}{30}$	$\frac{5}{30}$	$\frac{6}{30}$

(c) Find (i) E(Y)

(2)

(ii) Var(Y)

(3)

The random variable W = aX + b, where a and b are constants and a > 0Given that E(W) = E(Y) and Var(W) = Var(Y)

(d) find the value of a and the value of b.

**(5)** 

Each of the two spinners is spun once.

(e) Find P(W = Y)

**(2)** 

(a) if all the probabilities are the same, then we're looking at a discrete uniform distribution//

(b) (i) because this is only a 1 mark question, we are not expected to use the formula for expectation—

ue can therefore exploit symmetry:

X	10	12	14	16	] if this is a discrete uniform
P(X=3C)	P	P	P	P	_ ] if this is a discrete uniform distribution, the E(x) must be in the CENTRE



**Question 4 continued** 

$$\therefore E(\times) = 13$$

(ii) now we have to use the formula for variance:

n 4 part (b)(i)

$$= (10)^{2} + (12)^{2} + (14)^{2} + (16)^{2} - (13)^{2}$$

= 5

(c) now we're actually given the probability distribution in the form of a table, with the 'y' representing the options for the random variable Y to take and the P(Y=y) the likelihood

of each option to be taken

(i) Where, need to use the formula for expectation:

$$= 90/_{30} = 3$$
  
:. E(Y) = 3

$$\sqrt{(Y)} = (1)^{2} (\frac{4}{30}) + (2)^{2} (\frac{9}{30}) + (3)^{2} (\frac{6}{30}) + (4)^{2} (\frac{5}{30})$$

$$+ (5)^{2} (\frac{6}{30}) - (3)^{2}$$

```
Question 4 continued
```

(d) remembering that 
$$E(aW) = aE(W)$$
, and that  $E(W+X) = E(W) + E(X)$   
=) using  $W = aX + b$ 

$$E(w) = \alpha E(x) + b$$

sub into above:

we're also given that 
$$Y(W) = V(Y)$$

from part (c)(ii)

(and 
$$V(X) = 5$$

from part (a) (ii)

sub this into equtn 1:

$$13(0.6)+b=3$$

**Question 4 continued** 

(e) now that we have the values of 'a' and 'b', we can see that  $H = 0.6 \times -4.8$ 

from part (d)

where x tokes values: 10, 12, 14, 16

.. forming prob. distribution for W

$$W = 0.6(10)-4.8 \quad 0.6(12)-4.8 \quad 0.6(14)-4.8 \quad 0.6(16)-4.8$$
 $= 1.2 \quad = 2.4 \quad = 3.6 \quad = 4.8$ 

but these are all non-integers, and the y values are all integers .. P(W #Y)

∴0 values

Q4

(Total 16 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS ARE.

#### coefficient, linear regression, scatter diagrans

**5.** A company director wants to introduce a performance-related pay structure for her managers. A random sample of 15 managers is taken and the annual salary, *y* in £1000, was recorded for each manager. The director then calculated a performance score, *x*, for each of these managers.

The results are shown on the scatter diagram in Figure 1 on the next page.

(a) Describe the correlation between performance score and annual salary.

**(1)** 

The results are also summarised in the following statistics.

$$\sum x = 465$$
  $\sum y = 562$   $S_{xx} = 2492$   $\sum y^2 = 23140$   $\sum xy = 19428$ 

(b) (i) Show that  $S_{xy} = 2006$ 

**(1)** 

(ii) Find  $S_{yy}$ 

**(2)** 

(c) Find the product moment correlation coefficient between performance score and annual salary.

**(2)** 

The director believes that there is a linear relationship between performance score and annual salary.

(d) State, giving a reason, whether or not these data are consistent with the director's belief.

(1)

(e) Calculate the equation of the regression line of y on x, in the form y = a + bx Give the value of a and the value of b to 3 significant figures.

**(4)** 

(f) Give an interpretation of the value of b.

**(1)** 

(g) Plot your regression line on the scatter diagram in Figure 1

(2)

The director hears that one of the managers in the sample seems to be underperforming.

(h) On the scatter diagram, circle the point that best identifies this manager.

**(1)** 

The director decides to use this regression line for the new performance related pay structure.

(i) Estimate, to 3 significant figures, the new salary of a manager with a performance score of 30

**(2)** 

Leave blank

#### **Question 5 continued**

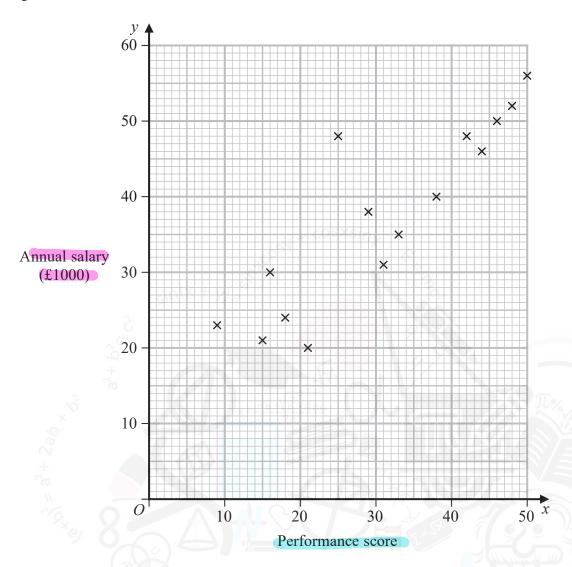


Figure 1

(a) we see a +ve correlation : as the performance score increases, the annual salary increases

(b) (i) He're asked to find the covariance between x and

sub into the above using info given

Turn over for a spare copy of the scatter diagram if you need to redraw your line.

**Question 5 continued** 

(ii) now we're asked to find the covariance between the 'y' variables

sub into the above

= 2083.7333...

$$= 2080(3 s.f)$$

(c) the PMCC measures the strength and type of relationship between the sc and the y

sub into above

= 0.8803104...

- (d) the r=0.880 is close to 1 : indicating a strong, the relationship, which is consistent with the director's belief
- (e) ne're asked to find the regression line of y on x-ne are given in the formula booklet that:

sub into the above

$$b = \frac{2006}{2492} = 0.80497...$$

... and the formula for 'a':

where 
$$\bar{y} = \frac{2y}{n}$$
 and  $\bar{x} = \frac{2x}{n}$ 

sub into the above:

$$y = \frac{562}{15}$$
 and  $x = \frac{465}{15} = 3$ 

Sub into formula for 'a':

$$\alpha = \frac{562}{15} - 0.80497...(31)$$

= 12.512413...

.. Subbing into general form: y=a+bx

(f) the value of 'b' represents the gradient of the regression line. however we need a worded explanation for it:

by £0.805 thousand i.e £800/

(9) He now need to plot y= 12.5+0.805x
4 remember key features:

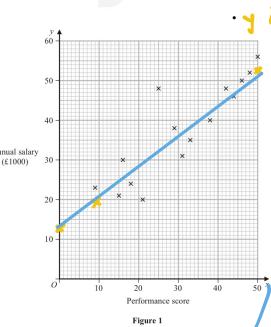
·x intercept = - 15.527...

·gradient = 0.805

upicking some points:

(h) underperforming => low performance score, high salary

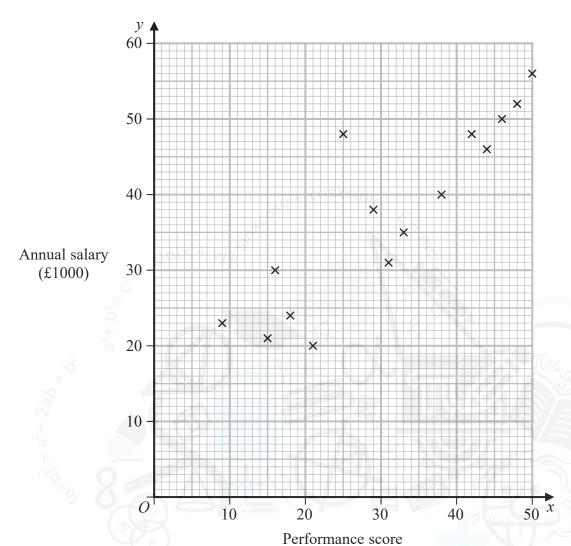
·· (25,48)



Leave blank

#### **Question 5 continued**

Only use this scatter diagram if you need to redraw your line.



y= 12.5+0.805 (30) =f 36.65 thousand

. £36,650

lor could've read off the regression line, although this is less accurate

(Total 17 marks)





Leave

blank

### Probability & Data Calculations - with geometry, standard deviation,

#### total mean and standard deviation

6. A disc of radius 1 cm is rolled onto a horizontal grid of rectangles so that the disc is equally likely to land anywhere on the grid. Each rectangle is 5 cm long and 3 cm wide. There are no gaps between the rectangles and the grid is sufficiently large so that no discs roll off the grid.

If the disc lands inside a rectangle without covering any part of the edges of the rectangle then a prize is won.

By considering the possible positions for the centre of the disc,

(a) show that the probability of winning a prize on any particular roll is  $\frac{1}{5}$ 

**(3)** 

A group of 15 students each roll the disc onto the grid twenty times and record the number of times, x, that each student wins a prize. Their results are summarised as follows

$$\sum x = 61 \qquad \sum x^2 = 295$$

(b) Find the standard deviation of the number of prizes won per student.

(2)

A second group of 12 students each roll the disc onto the grid twenty times and the mean number of prizes won per student is 3.5 with a standard deviation of 2

(c) Find the mean and standard deviation of the number of prizes won per student for the whole group of 27 students.

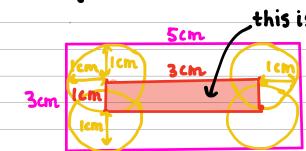
**(7)** 

The 27 students also recorded the number of times that the disc covered a corner of a rectangle and estimated the probability to be 0.2216 (to 4 decimal places).

(d) Explain how this probability could be used to find an estimate for the value of  $\pi$  and state the value of your estimate.

(3)

(a) the question is basically asking us to find where the centre of the disc can be such that the disc doesn't cover any of the edges of the rectangle—we think the the corners of the rectangle would be the boundaries and therefore anything inside it will mean a prize is WON



this is the area such that if centre of disc fell in it, a prize will be won

18



#### **Question 6 continued**

=> P(prize won)= 
$$\frac{1\times3}{3\times5}$$
 =  $\frac{1}{5}$  P(prize won)=  $\frac{1\times3}{3\times5}$  =  $\frac{1}{5}$ 

formula: 
$$V(x) = \frac{2x^2}{x^2} - (x)^2$$
, where  $x = \frac{2x}{x}$ 

sub into above

$$=\frac{295}{15}-(\frac{61}{15})^2$$

# (c) so we're being asked for the mean and s.d of all 27 students now

4 let the no. of prizes uon in 15 student group=x

4 let the no. of prizes uon in 12 student group = y

# where we can get $\frac{2}{3}$ by using the given into on $\frac{3}{3}$ = 3.5

Sub into above

```
Question 6 continued
```

# He can use the given info on s.d(Y)=2

# sub into above

$$2 = \sqrt{\frac{5y^2}{12} - (3.5)^2}$$

square both sides

$$4 = \frac{5y^2}{12} - 12.25$$

$$=) \frac{54^2}{12} = |6.25|$$

## sub into above.

(d) now we need to try to picture how the discs can cover the corners of the rectangle:

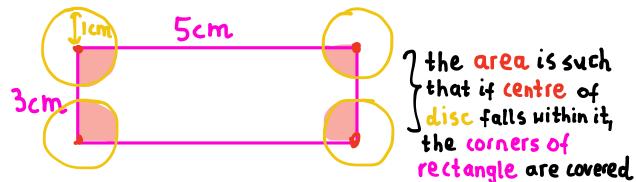
Q6

(Total 15 marks)

TOTAL FOR PAPER IS 75 MARKS

END





P(area) = 4 x area of quarter circles

area of rectangle

formula: mr2/4 = u(1)/4

sub into above:

 $= \frac{4 \times \pi/4}{15} = \frac{\pi}{15}$ 

we're given that Plarea) = 0.2216

=)  $0.2216 = \frac{\pi}{15}$ ×15 ×15

My Maths Cloud