Annotated by Tam HastingsMath.com

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Time 1 hour 30 minutes

Paper reference **WST01/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S1

You must have:
Mathematical Formulae and Statistical Tables (Yellow), calculator

Total Marks

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 7 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

L:1/1/1/

P 6 6 6 5 2 A 0 1 2 8

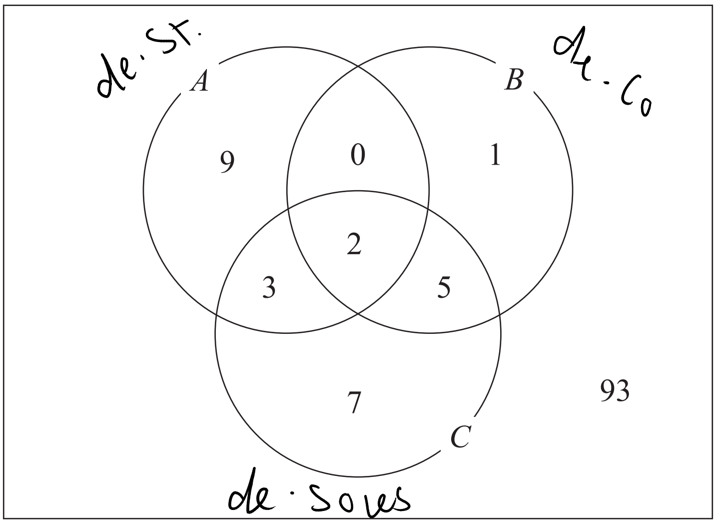
Pearson

Leave blank

1. A factory produces shoes.

A quality control inspector at the factory checks a sample of 120 shoes for each of three types of defect. The Venn diagram represents the inspector's results.

A represents the event that a shoe has defective stitching
B represents the event that a shoe has defective colouring
C represents the event that a shoe has defective soles



One of the shoes in the sample is selected at random.

(a) Find the probability that it does **not** have defective soles.
(1)

(b) Find $P(A \cap B \cap C')$
(1)

(c) Find $P(A \cup B \cup C')$
(2)

(d) Find the probability that the shoe has at most one type of defect.
(2)

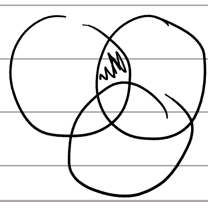
(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching.
(2)

The random variable X is the number of the events A, B, C that occur for a randomly selected shoe.

(f) Find $E(X)$ $\quad P(C) = 17$
(3)

(a) $\therefore P(C') = \frac{120-17}{120} = \frac{102}{120}$ ✓

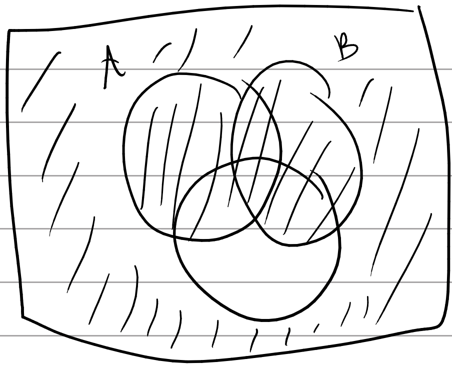
(b) $\frac{0}{120} = 0$ ✓ $\longrightarrow$



Question 1 continued

(c) $P(A \cup B \cup C')$

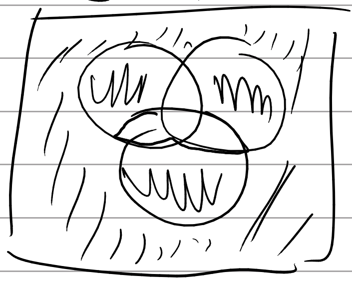
$\therefore = \dfrac{120-7}{120} = \dfrac{113}{120}$ ✓



(d) $P(\text{at Most one}) =$

$\dfrac{9+1+7+93}{120} = \dfrac{110}{120} = \dfrac{11}{12}$ ✓

zero / one



(e) $P(\underset{\text{shoe}}{\text{def}} \mid \underset{\text{at most}}{1\text{ def}})$

$= P(A \mid 1\text{ def at most}) = \dfrac{P(A \cap 1\text{ def at most})}{P(1\text{ def. at most})}$

$= \dfrac{\frac{9}{120}}{\frac{11}{12}}$

$= \dfrac{9}{110}$ ✓

(f) # of defects:

X	0	1	2	3
$P(X=x)$	$93/120$	$17/120$	$8/120$	$2/120$

$= 1$

(Probability dist. table)

$E(x) = \sum x\,P(x)$

$E(X) = 0 \times \dfrac{93}{120} + \ldots + 3 \times \dfrac{2}{120}$

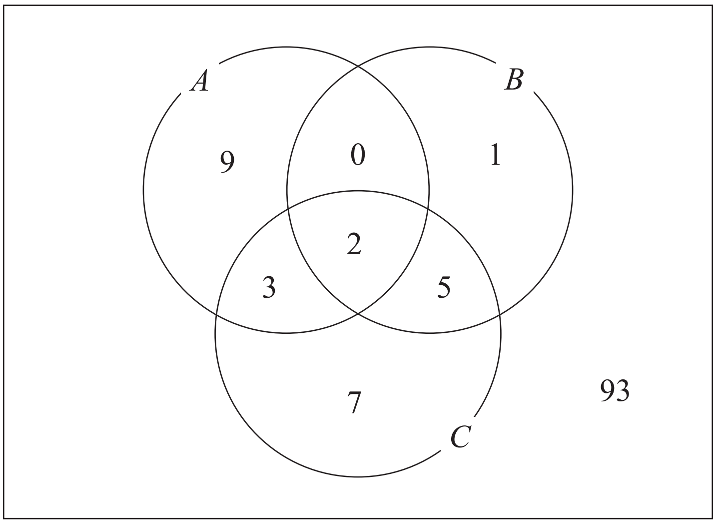
$= \dfrac{39}{120} = \dfrac{13}{40}$ ✓

Leave blank

Question 1 continued

Question 1 continued

This is a copy of the Venn diagram for this question.



(Total 11 marks)

Q1

Leave blank

Leave blank

2. Tom's car holds 50 litres of petrol when the fuel tank is full.

For each of 10 journeys, each starting with 50 litres of petrol in the fuel tank, Tom records the distance travelled, d kilometres, and the amount of petrol used, p litres.

The summary statistics for the 10 journeys are given below.

$n = 10$ $\quad \sum d = 1029 \quad \sum p = 50.8 \quad \sum dp = 5240.8 \quad S_{dd} = 344.9 \quad S_{pp} = 0.576$

(a) Calculate the product moment correlation coefficient between d and p **(3)**

The amount of petrol remaining in the fuel tank for each journey, w litres, is recorded.

(b) (i) Write down an equation for w in terms of p

(ii) Hence, write down the value of the product moment correlation coefficient between w and p **(2)**

(c) Write down the value of the product moment correlation coefficient between d and w **(1)**

a) $r = \dfrac{S_{xy}}{\sqrt{S_{xy} S_{yy}}} = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{(\Sigma(x-\bar{x})(\Sigma(y-\bar{y}))^2}}$

$= \dfrac{\Sigma xy - \dfrac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{\left(\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}\right)\left(\Sigma y^2 - \dfrac{(\Sigma y)^2}{n}\right)}}$

$r = \dfrac{S_{dp}}{\sqrt{S_{dd} S_{pp}}} = \dfrac{5240.8 - \dfrac{1029 \times 50.8}{10}}{\sqrt{344.9 \times 0.576}}$

$= 0.95638$

$= 0.956$ ✓

P 6 6 6 5 2 A 0 6 2 8

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

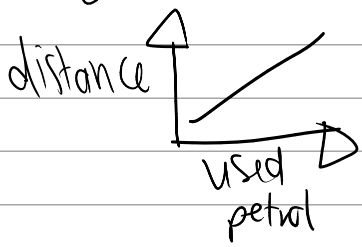
Question 2 continued

(b) (i) $w = 50 - p$ ✓

amount remain = amount start − amount use

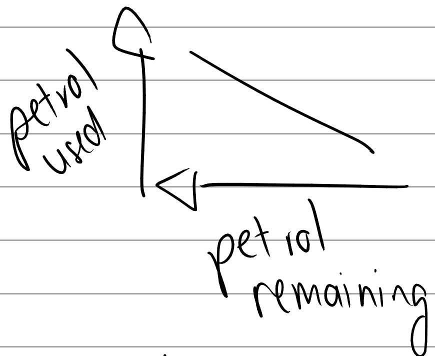
(ii)

↳ r between petrol used and distance = strong positive correlation



∴ but for petrol remaining vs petrol used
(w) (p)

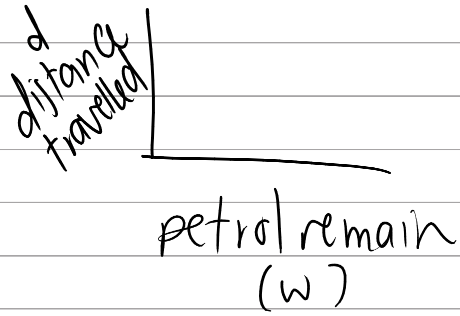
as one goes down the other goes up in exact proportion



∴ $r = -1$

perfect negative correlation

(c) r for d and w



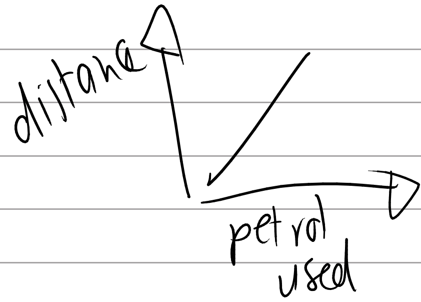
Turn over

P 6 6 6 5 2 A 0 7 2 8

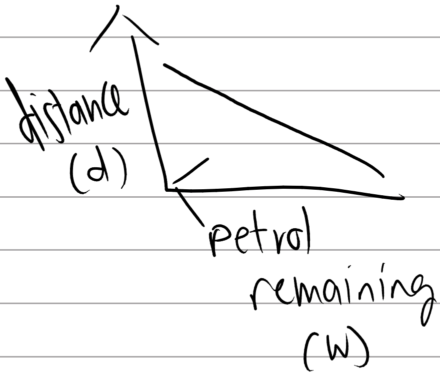
Leave blank

Question 2 continued

this is the opposite of



it is



but in exact proportion

$\therefore r = -0.956$ ✓ → from a

Leave blank

Question 2 continued

Q2

(Total 6 marks)

Leave blank

3. The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

$n=24$

Key: 10 | 8 represents 108 deliveries

10	8	9										(2)
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	a	b	c									(3)

where a, b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than $1.5 \times$ interquartile range above the upper quartile.

Given that there is only one outlier for these data, c is outlier

(a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained → var & SD. not affected by + −

Coding:

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

(b) Find the mean number of deliveries.

(3)

(c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

(d) Find $\mathrm{P}(D > 118 | X < 0)$

(2)

9 RTP: c = 9

Outlier = 1.5 × IQR + Q3

Question 3 continued

$IQR = UQ - LQ$

position Q_1: $\frac{n}{4} = \frac{24}{4} = 6$ $\therefore \underbrace{6^{th} \& 7^{th}}$ (even #)

$\therefore Q_1 = \frac{116+116}{2} = 116$

position $Q_3 = \frac{3n}{4} = \frac{3(24)}{4} = 18$ $\therefore 18^{th} \& 19^{th}$ (even #)

$\therefore Q_3 = \frac{125+125}{2} = 125$

$\therefore IQR = 125 - 116 = 9$

$\therefore$ outlier $= 1.5(9) + 125$

$= 138.5$ $= 139$ (rounded)

Key 13|9

$\therefore c = 9$

(b) coding

$x = d - 125$ $\Sigma x = -96$ $\Sigma(x-\bar{x})^2 = 1306$

$n = 24$

RTP: mean of deliveries

$\bar{x} = \frac{\Sigma x}{n} = \frac{-96}{24} = -4$ (coded mean

$\bar{x} = \bar{d} - 125$ (mean is affected in coding)

$\bar{d} = \bar{x} + 125$

$\bar{d} = -4 + 125 = 121$ ✓

Question 3 continued

(c) RTP: S.D $d = ?$

$\sigma = \sqrt{MS - SM}$

$\sigma_x = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\dfrac{1306}{24}}$

$= 7.3767$

S.D not affected by $+/-$; only by $\times/\div$

$\therefore \sigma_d = 7.3767$
$= 7.38$ (3sf) ✓

(d) RTP: $P(D>118 \mid X<0)$ (Conditional probability)

change $X<0$ in terms of D

$X = D - 125$

$\therefore D - 125 < 0$

$\therefore D < 125$

$\therefore$ RTP: $P(D>118 \mid D<125)$

$= \dfrac{P((D>118) \cap (D<125))}{P(D<125)}$

Question 3 continued

$P((O > 118 \cap O < 125)$

is from 119 to 124

$\therefore$ 5 entries

$\therefore \frac{5}{24}$

$P(O < 125)$ is 14 entries

$\therefore \frac{14}{24}$

$\therefore \dfrac{\frac{5}{24}}{\frac{14}{24}}$

$= \frac{5}{14}$ ✓

(Total 10 marks)

Leave blank

Q3

discrete random variables

4. The random variable W has a discrete uniform distribution where

$$P(W = w) = \frac{1}{5} \quad \text{for } w = 1, 2, 3, 4, 5$$

(a) Find $P(2 \leqslant W < 3.5)$ **(1)**

The discrete random variable $X = 5 - 2W$

(b) Find $E(X)$ **(3)**

(c) Find $P(X < W)$ **(2)**

The discrete random variable $Y = \frac{1}{W}$

(d) Find

(i) the probability distribution of Y

(ii) $\text{Var}(Y)$, showing your working. **(5)**

(e) Find $\text{Var}(2 - 3Y)$ **(2)**

(a) RTP: $P(2 \leq W < 3.5)$

because discrete

$= P(2 \leq W < 3)$

$\therefore P(W=2) + P(W=3) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$ ✓

(b) $E(x) = ?$ if $X = 5 - 2W$

$\therefore E(5 - 2W)$

coded mean is affected by $\times \div$ and $+ -$

$\therefore -2[E(W)] + 5 = E(X)$

Question 4 continued

$E(W) = \frac{n+1}{2}$

$= 3$

can also use
$(1 \times \frac{1}{5}) + (2 \times \frac{1}{5}) + (3 \times \frac{1}{5}) \ldots$
to find $E(X)$

$\therefore E(X) = -2(3) + 5$

$= -1$ ✓

(c) RTP: $P(X < W)$

rewrite in terms of W:

$X = 5 - 2W$

$\therefore P(5 - 2W < W)$

$\therefore P(5 < 3W)$

$\therefore P(\frac{5}{3} < W)$

$\therefore P(W > \frac{5}{3})$

$\therefore P(W > 1\frac{1}{3})$

$\therefore P(W \geq 2)$ because discrete

$= 4 \times \frac{1}{5} = \frac{4}{5}$ ✓

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Leave blank

Question 4 continued

d) (i) $Y = \frac{1}{W}$

w	1	2	3	4	5
y	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$
$P(Y=y)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
y^2	1	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{1}{16}$	$\frac{1}{25}$

→ (uniform distribution)

(ii) $Var(Y) = ?$

$Var = MS - SM$

$= E(Y^2) - [E(Y)]^2$

$E(Y) = (1 \times \frac{1}{5}) + (\frac{1}{2} \times \frac{1}{5}) + \ldots + (\frac{1}{5} \times \frac{1}{5})$

$= \frac{137}{300}$

$E(Y^2) = (1 + \frac{1}{5}) + (\frac{1}{4} \times \frac{1}{5}) + \ldots + (\frac{1}{25} \times \frac{1}{5})$

$= \frac{5269}{18000}$

$\therefore Var(Y) = \frac{5269}{18000} - \left(\frac{137}{300}\right)^2$

$= \frac{947}{11250} = 0.0842$ ✓

Leave blank

Question 4 continued

(e) $\mathrm{Var}(2-3Y) = ?$

$= (-3)^2(\mathrm{Var}(Y))$

$= 9 \times \frac{947}{11250}$

$= \frac{947}{1250}$

$= 0.7576$ ✓

CODING

var(x) is affected by $\times \div$

but NOT by $+ -$

also square it

Q4

(Total 13 marks)

Normal distribution

5. Jia writes a computer program that randomly generates values from a normal distribution.

He sets the mean as 40 and the standard deviation as 2.4 $\quad V \sim N(40, 24^2)$

(a) Find the probability that a particular value generated by the computer program is less than 37 $\quad P(X < 37)$

(3)

Jia changes the mean to m but leaves the standard deviation as 2.4 $\quad \to P(A) \times P(B) = P(A \cap B)$

The computer program then randomly generates 2 independent values from this normal distribution.

The probability that both of these values are greater than 32 is 0.16 $\quad P(X > 32) = 0.16$

(b) Find the value of m, giving your answer to 2 decimal places. $\quad m = ?$

(4)

Jia now changes the mean to 4 and the standard deviation to 8 $\quad \to P(A) \times P(B) \times \ldots$

The computer program then randomly generates 5 independent values from this normal distribution.

(c) Find the probability that at least one of these values is negative. $\quad P(\text{at least 1 neg})$

(4)

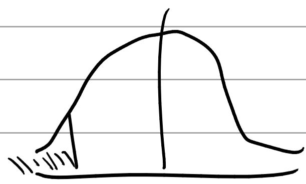
(a) $\mu = 40 \quad S.D = 2.4$

$V \sim N(40; 2.4^2)$

$P(V < 37)$ need to standardise $Z = \dfrac{x - \mu}{\sigma}$

$= P\left(Z < \dfrac{37-40}{2.4}\right)$

$= P(Z < -1.25)$

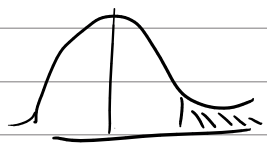


V $\quad$ 37 $\quad$ 40

Z $\quad$ −1.25 $\quad$ 0

tables ∴

$= P(Z > 1.25)$ $\quad \to$ use symmetry

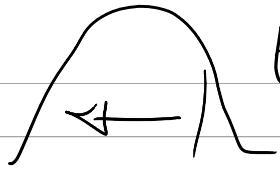


Z $\quad$ 1.25

Leave blank

Question 5 continued

$= 1 - P(Z < 1.25)$



tables go this way

$= 1 - 0.8944$ → from tables

$= 0.1056$

$= 0.106$ (3sf) ✓

(b) $V \sim N(m; 24^2)$

independent events

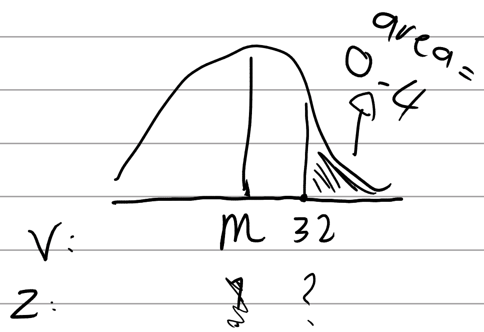
$P(V > 32) \times P(V > 32) = 0.16$

$P(A) \times P(B) = 0.16$

$= [P(V > 32)]^2 = 0.16$

$= P(V > 32) = \sqrt{0.16} = 0.4$

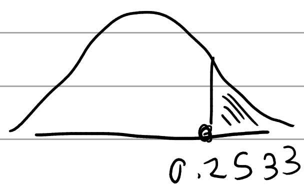
need to standardise



use table with areas to the right of z-values (small table)

p	z
0.400	0.2533

$\therefore P(V > 32) = 0.4$



Question 5 continued

$$z = \frac{32 - m}{2.4}$$

$$0.2533 = \frac{32 - m}{2.4}$$

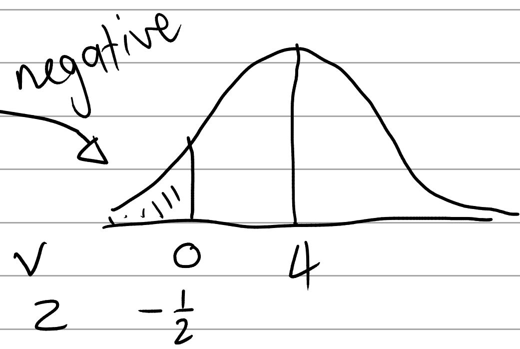
$$m = 31.392$$
$$= 31.4 \quad (3sf) \checkmark$$

(c) $V \sim N(4, 8^2)$

$P(V < 0)$

$= P\left(Z < \frac{0-4}{8}\right)$

$= P\left(Z < -\frac{1}{2}\right)$



if at least one is negative, then
1 neg 4 pos ✓
2 neg 3 pos ✓
3 neg ~~1~~ 2 pos ✓
4 neg 1 pos ✓
5 neg 0 pos ✓
0 neg 5 pos ✗

work out P(all five pos) and minus from 1

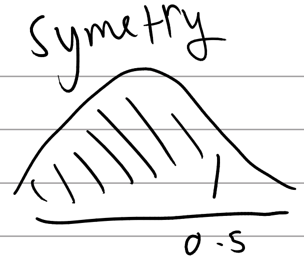
$\therefore \; 1 - [P(V > 0)]^5$

all five positive

Leave blank

Question 5 continued

$\therefore \ 1 - [P(z < 0.5)]^5$



z

$= 1 - (0.6915)^5$ from tables

$= 0.841889$

$= 0.842 \quad (3sf)$ ✓

Q5

(Total 11 marks)

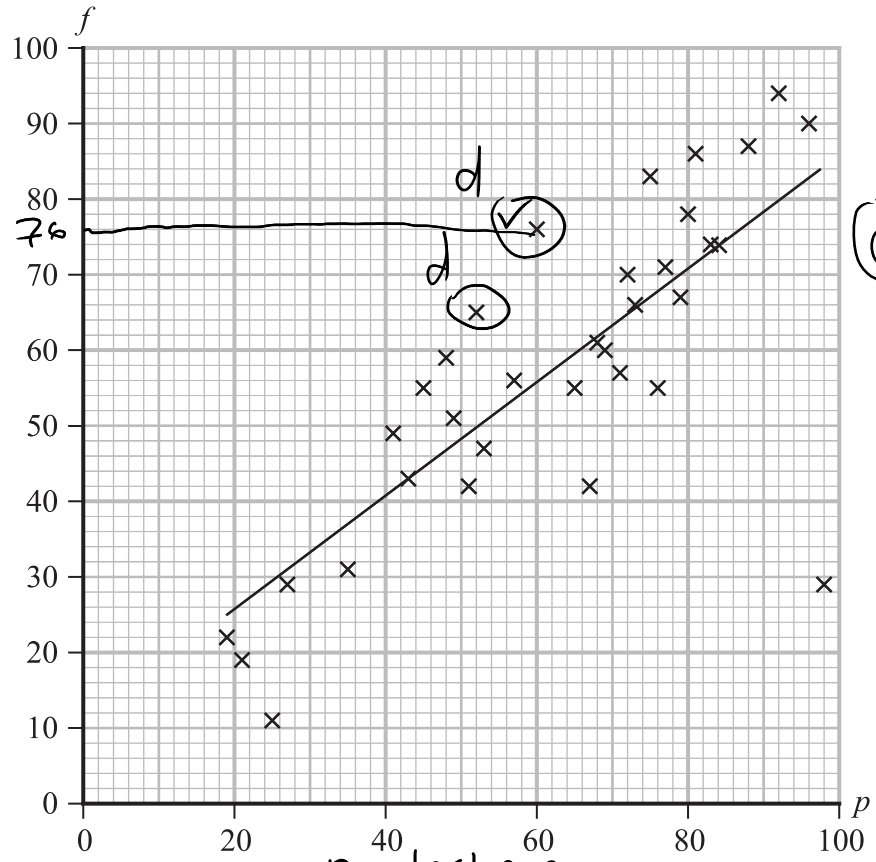
Correlation / Regression analysis

Leave blank

6. Students on a psychology course were given a pre-test at the start of the course and a final exam at the end of the course. The teacher recorded the number of marks achieved on the pre-test, p, and the number of marks achieved on the final exam, f, for 34 students and displayed them on the scatter diagram.

$n=34$

final exam (f)



pretest (p)

(d) lower on p; higher on f

The equation of the least squares regression line for these data is found to be

$y=ax+b$

$$f = 10.8 + 0.748p$$

For these students, the mean number of marks on the pre-test is 62.4 (p)

(a) Use the regression model to find the mean number of marks on the final exam. (f) **(2)**

(b) Give an interpretation of the gradient of the regression line. $m = \frac{0.748}{1}$ **(1)**

Considering the equation of the regression line, Priya says that she would expect someone who scored 0 marks on the pre-test to score 10.8 marks on the final exam.

(p) (f)

(c) Comment on the reliability of Priya's statement. ? **(1)**

(f)

(d) Write down the number of marks achieved on the final exam for the student who exceeded the expectation of the regression model by the largest number of marks. **(1)**

Question 6 continues on page 24.

Question 6 continued

(a) sub $p = 62.4$

$\therefore \quad f = 10.8 + 0.748(62.4)$

mean marks (f) $= 57.5$ (3sf) ✓

(b) $m_f = \frac{0.748}{1} = \frac{\Delta y}{\Delta x}$

positive gradient

for every increase of one mark in the pretest, the final exam increases by 0.748

(c) not reliable → data values do not have zero on the pretest.
Therefore unreliable as this would be extrapolation as it is out of the range of the data values.

(d) exceed regression model : maximum difference

$\therefore f - p$ (by inspection) → lower on p
→ higher on f

$76 - 67 = 16$ ✓
$65 - 52 = 13$

Question 6 continues on page 24.

$\therefore 76$ marks ✓

Question 6 continued

(e) Find the range of values of p for which this regression model, $f = 10.8 + 0.748p$, predicts a greater number of marks on the final exam than on the pre-test.

(f) (p)

(3)

Later the teacher discovers an error in the recorded data. The student who achieved a score of 98 on the pre-test, scored 92 not 29 on the final exam.

(p) (f)

The summary statistics used for the model $f = 10.8 + 0.748p$ are corrected to include this information and a new least squares regression line is found.

Given the **original** summary statistics were,

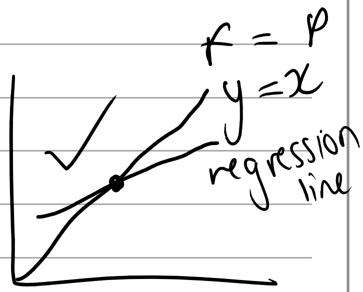
$$n = 34 \quad \sum p = 2120 \quad \sum pf = 133\,486 \quad S_{pp} = 15\,573.76 \quad S_{pf} = 11\,648.35$$

$y = 5x + 9$

(f) calculate the gradient of the new regression line. Show your working clearly. $a = ?$

(5)

(e) $y = x$ line is where $(p) = (f)$



choose values above the line
$f = p$

find where regression line intersects $f = p$

$f = 10.8 + 0.748p$

sub $f = p$

$p = 10.8 + 0.748p$

$p - 0.748p = 10.8$

$0.252p = 10.8$

$p = \frac{10.8}{0.252}$

$= 42.857$

$= 42.9$ (3sf)

Question 6 continued

When regression line is above $f = p$ we have

p greater than f

$\therefore$ p greater than point of intersection

$p < 42.9$ (3sf) ✓

(f) gradient of regression line $= b$

$b = \frac{S_{xy}}{S_{xx}}$ $\quad S_{xy} = \Sigma(x-\bar{x})(y-\bar{y})$

$= \Sigma xy - \frac{\Sigma x \Sigma y}{n}$

$\therefore b = \frac{S_{pf}}{S_{pp}}$

RTP: new S_{pf} and new S_{pp}

the mistake was in the final exam (f) $\to$ given

$\therefore$ new S_{pp} not affected $= 15573.76 \to$ given

S_{pf} is affected $= \Sigma pf - \frac{\Sigma p \Sigma f}{n}$ $\to$ anything with f is affected

new $\Sigma pf = ?$ new $\Sigma f = ?$

old S_{pf} = old Σpf $-$ $\frac{\text{old } \Sigma p \times (\text{old } \Sigma f)}{34}$

$11.6448.35 = 133486 - \frac{2120 \times \text{old } \Sigma f}{24}$

old $\Sigma f = 1954$

new $\Sigma f = 1945 - 29 + 92 = 2017$

(Total 13 marks)

Q6

Discrete random variables

7. A bag contains n marbles of which 7 are green.

From the bag, 3 marbles are selected at random.

The random variable X represents the number of green marbles selected. [X is green marbles]

The cumulative distribution function of X is given by

	x	0	1	2	3	
cum. fr:	$\mathrm{F}(x)$	a	b	$\frac{37}{38}$	1	$0.1 - \frac{37}{38}$
	$P(X=x)$	a	$b-a$	$\frac{37}{38}-(b-a)$	$\frac{1}{38}$	

(a) Show that $n(n-1)(n-2) = 7980$

(4)

(b) Verify that $n = 21$ satisfies the equation in part (a).

(1)

Given that $n = 21$

(c) find the exact value of a and the exact value of b

(6)

6(f) continued

$$\text{old } \Sigma_{pf} = 133486$$

given

$$\text{New } \Sigma_{pf} = \text{old } \Sigma_{pf} - (98 \times 29) + (98 \times 92)$$

remove old sum pf mistake; add new sum pf fixed

$$= 139660$$

$$\therefore \text{new } S_{pf} = 139660 - \frac{2120 \times 2017}{34}$$

$$= 13894.12$$

$$\therefore b = \frac{13894.12}{15573.76} \qquad b = \frac{S_{pf}}{S_{pp}}$$

$$= 0.892$$

start here

Question 7 ~~continued~~

7 a

$p(3\text{green marbles})$

$$P(X=3) = F(3) - F(2) = \frac{1}{38}$$

$$\frac{7}{n} \times \frac{6}{n-1} \times \frac{5}{n-2} = \frac{1}{38}$$

→ green marbles picked ∴ one less marble in the bag each pick

$$7 \times 6 \times 5 \times 38 = n(n-1)(n-2)$$

$$n(n-1)(n-2) = 7980 \checkmark$$

b) $n(n-1)(n-2)$

$$21(20)(19) = 7980 \checkmark$$

$\therefore n = 21$ satisfies the equation

c) $a = ?$ $b = ?$

green $= 7$
not green $= 21 - 7 = 14$

a: $p(X=0)$ → no green marbles

Not G × Not G × Not G

$$a = \frac{14}{21} \times \frac{13}{20} \times \frac{12}{19} = \frac{26}{95} \checkmark$$

b: $b = F(1)$

$$= P(X=0) + P(X=1)$$

→ arranged 3 ways

$$P(X=1) = 3\big(P(G) \times P(\text{Not G}) \times P(\text{Not G})\big)$$

$$= 3\left(\frac{14}{21} \times \frac{13}{20} \times \frac{7}{19}\right)$$

$$= \frac{91}{190}$$

Leave blank

Leave blank

Question 7 continued

$F(1) = P(X=0) + P(X=1)$

$= \frac{26}{95} + \frac{91}{190}$

$= \frac{143}{190}$

$a = \frac{26}{95}$ $b = \frac{143}{190}$

Q7

(Total 11 marks)

TOTAL FOR PAPER: 75 MARKS

END