Write your name here

Surname	Other names
aymanzayedmannan	

Pearson Edexcel
International Advanced Level

Centre Number	Candidate Number

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 27 January 2016 – Morning
Time: 1 hour 30 minutes

Paper Reference
WST01/01

You must have:
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

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 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). Coloured pencils and highlighter pens must not be used.
- **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. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- 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.

Turn over ▶

P46949A

1/1/1/1/1/

P 4 6 9 4 9 A 0 1 2 4

PEARSON

1. The discrete random variable X has the probability distribution given in the table below.

x	-2	1	3	4	6
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{6}$

(a) Write down the value of F(5)

(1)

(b) Find E(X)

(2)

(c) Find Var(X)

(3)

The random variable $Y = 7 - 2X$

(d) Find

(i) E(Y)

(ii) Var(Y)

(iii) P($Y > X$)

(6)

1(a) $F(5) = P(X=-2) + P(X=1) + P(X=3) + P(X=4)$

$= \frac{1}{4} + \frac{1}{6} + \frac{1}{3} + \frac{1}{12}$

$\therefore F(5) = \frac{5}{6}$

(b) $E(X) = -2(\frac{1}{4}) + 1(\frac{1}{6}) + 3(\frac{1}{3}) + 4(\frac{1}{12}) + 6(\frac{1}{6})$

$\therefore E(X) = 2$

(c) $E(X^2) = (-2)^2(\frac{1}{4}) + 1^2(\frac{1}{6}) + 3^2(\frac{1}{3}) + 4^2(\frac{1}{12}) + 6^2(\frac{1}{6})$

$= \frac{23}{2}$

$Var(X) = E(X^2) - [E(X)]^2$

$= (\frac{23}{2})^2 - 2^2$

$\therefore Var(X) = \frac{15}{2}$

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

P 4 6 9 4 9 A 0 2 2 4

Question 1 continued

(d)(i) $E(Y) = E(7-2X)$

$= 7 - 2E(X)$

$= 7 - 2(2)$

$= 7 - 4$

$\therefore E(Y) = 3$

(ii) $Var(Y) = Var(7-2X)$

$= 2^2 \cdot Var(X)$

$= 4Var(X)$

$= 4 \times \frac{15}{2}$

$\therefore Var(Y) = 30$

(iii) $P(Y > X)$

$= P(7-2X > X)$

$= P(-3X > -7)$

$= P(3X < 7)$

$= P(X < 2\frac{1}{3})$

$= P(X \leq 2)$

$= P(X = -2) + P(X = 1)$

$= \frac{1}{4} + \frac{1}{6}$

$\therefore P(Y > X) = \frac{5}{12}$

Q1

(Total 12 marks)

Turn over

2.

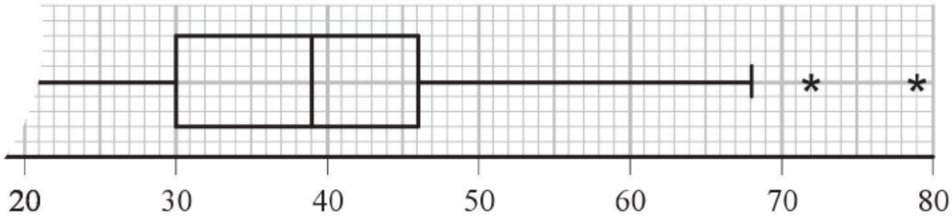


Figure 1

Figure 1 shows part of a box and whisker plot for the marks in an examination with a large number of candidates. Part of the lower whisker has been torn off.

(a) Given that 75% of the candidates passed the examination, state the lowest mark for the award of a pass.

(1)

(b) Given that the top 25% of the candidates achieved a merit grade, state the lowest mark for the award of a merit grade.

(1)

An outlier is defined as any value greater than c or any value less than d where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(c) Find the value of c and the value of d.

(2)

(d) Write down the 3 highest marks scored in the examination.

(2)

The 3 lowest marks in the examination were 5, 10 and 15

(e) On the diagram on page 7, complete the box and whisker plot.

(3)

Three candidates are selected at random from those who took this examination.

(f) Find the probability that all 3 of these candidates passed the examination but only 2 achieved a merit grade.

(3)

2(a) Lowest pass mark: 30

(b) Lowest pass mark for a merit: 46

Question 2 continued

(c) $c = Q_3 + 1.5(Q_3 - Q_1)$

$= 46 + 1.5(46 - 30)$

$\therefore c = 70$

$d = Q_1 - 1.5(Q_3 - Q_1)$

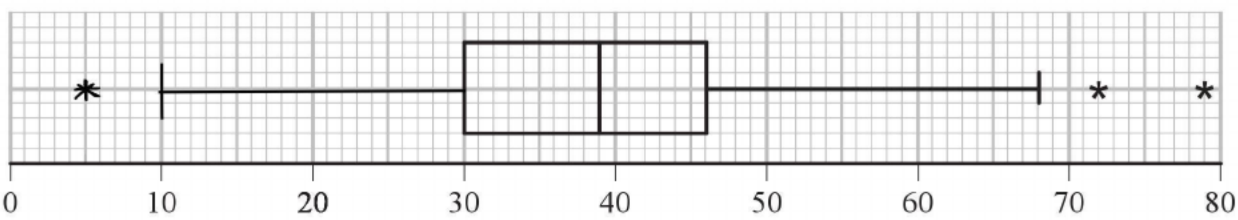
$= 30 - 1.5(46 - 30)$

$\therefore d = 6$

(d) The three highest scores: 68; 72; 79.

(f) $P(\text{3 passes, only two merits}) = 3[(0.25)(0.25)(0.75 - 0.25)]$

$= \frac{3}{32}$



Turn over for a spare diagram if you need to redraw your plot.

3. A publisher collects information about the amount spent on advertising, £ x, and the sales, y books, for some of her publications. She collects information for a random sample of 8 textbooks and codes the data using $v = \frac{x+50}{200}$ and $s = \frac{y}{1000}$ to give

v	0.60	8.10	4.30	0.40	1.60	6.40	2.50	5.10
s	1.84	6.73	5.95	1.30	2.45	7.46	4.82	6.25

[You may use: $\sum v = 29$ $\sum s = 36.8$ $\sum s^2 = 209.72$ $\sum vs = 177.311$ $S_{vv} = 55.275$]

(a) Find S_{vs} and S_{ss}

(3)

(b) Calculate the product moment correlation coefficient for these data.

(2)

The publisher believes that a linear regression model may be appropriate to describe these data.

(c) State, giving a reason, whether or not your answer to part (b) supports the publisher's belief.

(1)

(d) Find the equation of the regression line of s on v, giving your answer in the form $s = a + bv$

(4)

(e) Hence find the equation of the regression line of y on x for the sample of textbooks, giving your answer in the form $y = c + dx$

(3)

The publisher calculated the regression line for a sample of novels and obtained the equation

$$y = 3100 + 1.2x$$

She wants to increase the sales of books by spending more money on advertising.

(f) State, giving your reasons, whether the publisher should spend more money on advertising textbooks or novels.

(2)

3(a) $S_{vs} = \sum vs - \frac{\sum v \sum s}{n}$

$= 177.311 - \frac{29(36.8)}{8}$

$\therefore S_{vs} = 43.911$

DO NOT WRITE IN THIS AREA

Question 3 continued

$$S_{ss} = \sum s^2 - \frac{\sum s \sum s}{n}$$

$$= 209.72 - \frac{36.8^2}{8}$$

$$\therefore S_{ss} = 40.44$$

(b) $$r = \frac{S_{vs}}{\sqrt{S_{ss} S_{vv}}}$$

$$= \frac{43.911}{\sqrt{40.44(55.275)}}$$

$$= 0.928759\ldots$$

$$\therefore r \approx 0.929 \quad (3\text{ SF})$$

(d) $$b = \frac{S_{vs}}{S_{vv}}$$

$$= \frac{43.911}{55.275}$$

$$= 0.794409\ldots \simeq 0.794 \quad (3\text{ SF})$$

$$a = \bar{s} - b\bar{v}$$

$$= \frac{\sum s}{n} - b \cdot \frac{\sum v}{n}$$

$$= \frac{36.8}{8} - (0.794\ldots)\left(\frac{29}{8}\right)$$

$$= 1.7202645\ldots \approx 1.72 \quad (3\text{ SF})$$

Regression line:

$$s = 1.72 + 0.794v \quad (3\text{ SF})$$

(e) $$\frac{y}{1000} = 1.72 + 0.794\left(\frac{x+50}{200}\right)$$

$$y = 1720 + 794\left(\frac{x+50}{200}\right)$$

$$= 1720 + 3.97x + 198.5$$

$$= 1918.5 + 3.97x$$

$$\therefore y = 1920 + 3.97x \quad (3\text{ SF})$$

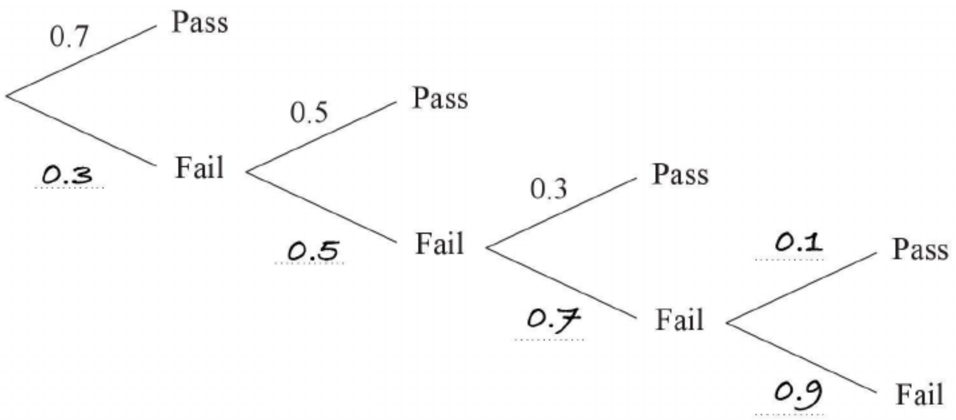
Q3

(Total 15 marks)

P46949A01324

4. A training agency awards a certificate to each student who passes a test while completing a course.
Students failing the test will attempt the test again up to 3 more times, and, if they pass the test, will be awarded a certificate.
The probability of passing the test at the first attempt is 0.7, but the probability of passing reduces by 0.2 at each attempt.

(a) Complete the tree diagram below to show this information.



(2)

A student who completed the course is selected at random.

(b) Find the probability that the student was awarded a certificate.

(2)

(c) Given that the student was awarded a certificate, find the probability that the student passed on the first or second attempt.

(3)

The training agency decides to alter the test taken by the students while completing the course, but will not allow more than 2 attempts. The agency requires the probability of passing the test at the first attempt to be p, and the probability of passing the test at the second attempt to be $(p - 0.2)$. The percentage of students who complete the course and are awarded a certificate is to be 95%

(d) Show that p satisfies the equation

$$p^2 - 2.2p + 1.15 = 0$$

(3)

(e) Hence find the value of p, giving your answer to 3 decimal places.

(3)

4(b) $P(\text{pass}) = 0.7 + (0.3)(0.5) + 0.3(05)(0.3) + 0.3(05)(0.7)(0.1)$

$= 0.7 + 0.15 + 0.045 + 0.0105$

$\therefore P(\text{pass}) = 0.9055$

(c) $P(\text{pass on 1st or 2nd attempt} \mid \text{pass}) = \dfrac{P(\text{pass on 1st or 2nd attempt} \cap \text{pass})}{P(\text{pass})}$

$= \dfrac{0.7 + 0.3(0.5)}{0.9055}$

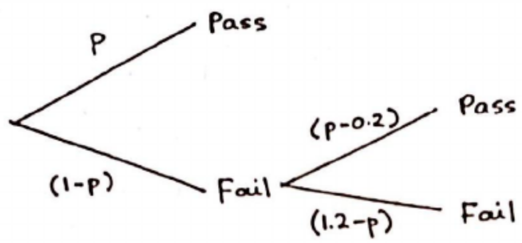
$= \dfrac{0.85}{0.9055}$

$= \dfrac{1700}{1811}$

$= 0.938707\ldots$

≈ 0.9387 (4 DP)

(d)



$P(\text{pass}) = p + (1-p)(p-0.2) = 95\%$

$\Rightarrow p + p - 0.2 - p^2 + 0.2p = 0.95$

$-p^2 + 2.2p - 0.2 - 0.95 = 0$

$-p^2 + 2.2p - 1.15 = 0$

$\therefore p^2 - 2.2p + 1.15 = 0$ QED

(e) $p = \dfrac{-(-2.2) \pm \sqrt{(-2.2)^2 - 4(1)(1.15)}}{2(1)}$

$= \dfrac{2.2 \pm \sqrt{6}/5}{2}$

$= 1.1 \pm \dfrac{\sqrt{6}}{10}$

$\therefore p = 0.855051\ldots$ or $p = 1.3449\ldots$

$\therefore p \approx 0.855$ (3 DP) $\left[\because 0 \le p \le 1 \Leftrightarrow p \ne 1.1 + \dfrac{\sqrt{6}}{10}\right]$

Q4

(Total 13 marks)

5. Rosie keeps bees. The amount of honey, in kg, produced by a hive of Rosie's bees in a season, is modelled by a normal distribution with a mean of 22 kg and a standard deviation of 10 kg.

(a) Find the probability that a hive of Rosie's bees produces less than 18 kg of honey in a season.

(3)

The local bee keepers' club awards a certificate to every hive that produces more than 39 kg of honey in a season, and a medal to every hive that produces more than 50 kg in a season. Given that one of Rosie's bee hives is awarded a certificate

(b) find the probability that this hive is also awarded a medal.

(5)

Sam also keeps bees. The amount of honey, in kg, produced by a hive of Sam's bees in a season, is modelled by a normal distribution with mean μ kg and standard deviation σ kg. The probability that a hive of Sam's bees produces less than 28 kg of honey in a season is 0.8413

Only 20% of Sam's bee hives produce less than 18 kg of honey in a season.

(c) Find the value of μ and the value of σ. Give your answers to 2 decimal places.

(6)

5. let X be the amount of honey, in kg, produced by a hive of Rosie's bees

$X \sim N(22, 10^2)$

(a) $P(X < 18)$

$= P\left(Z < \frac{18-22}{10}\right)$

$= P(Z < -0.40)$

$= P(Z > 0.40)$

$= 1 - P(Z < 0.40)$

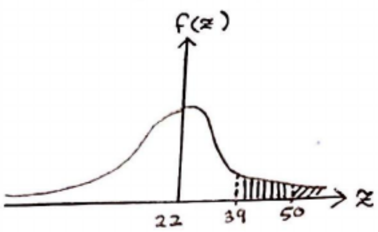
$= 1 - \Phi(0.40)$

$= 1 - 0.6554$ (tables)

$= 0.3446$

(b) $P(\text{certificate}) = P(X > 39)$

$P(\text{medal}) = P(X > 50)$



$P(\text{medal} \mid \text{certificate}) = \frac{P(\text{medal} \cap \text{certificate})}{P(\text{certificate})}$

$= \frac{P(X > 50)}{P(X > 39)}$

Question 5 continued

$$= \frac{1 - P\left(Z < \frac{50-22}{10}\right)}{1 - P\left(Z < \frac{39-22}{10}\right)}$$

$$= \frac{1 - \Phi(2.80)}{1 - \Phi(1.70)}$$

$$= \frac{1 - 0.9974}{1 - 0.9554} \quad \text{(tables)}$$

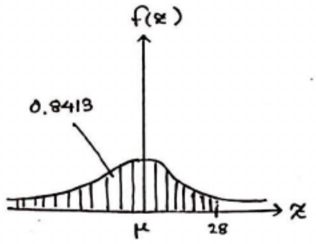
$$= \frac{13}{233}$$

$$= 0.058295\ldots$$

$\therefore P(\text{medal} \mid \text{certificate}) \approx 0.0583$ (4 DP)

(c) Let Y be the amount of honey, in kg, produced by a hive of Sam's bees.

$\therefore Y \sim N(\mu, \sigma^2)$

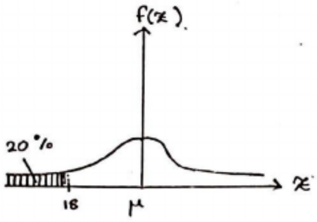


$P(Y < 28) = 0.8413$

$P\left(Z < \frac{28-\mu}{\sigma}\right) = 0.8413$

$\frac{28-\mu}{\sigma} = 1.00$ (tables)

$28 - \mu = \sigma$ ——— (i)



$P(Y < \cancel{20}\ 18) = 20\%$

$P\left(Z < \frac{\cancel{20}\ 18 - \mu}{\sigma}\right) = 0.20$

$\frac{\cancel{20}\ 18 - \mu}{\sigma} = -0.8416$ (tables)

$18\ \cancel{20} - \mu = -0.8416\sigma$ ——— (ii)

(i) ⇒ $28 - \mu = \sigma$

(ii) ⇒ $18 - \mu = -0.8416\sigma$

(−) $10 = 1.8416\sigma$

$\sigma = \frac{6250}{1151} = 5.43006\ldots \approx 5.43$ kg (3 SF)

Sub. σ in (i)

$28 - \mu = \frac{6250}{1151}$

$-\mu = \frac{6250}{1151} - 28$

$\mu = 22.5699\ldots \approx 22.6$ kg (3 SF)

$\therefore Y \sim N(22.6, 5.43^2)$ (3 SF)

Q5

(Total 14 marks)

6. Yujie is investigating the weights of 10 young rabbits. She records the weight, x grams, of each rabbit and the results are summarised below.

$$\sum x = 8360 \quad \text{and} \quad \sum (x - \bar{x})^2 = 63\,840$$

(a) Calculate the mean and the standard deviation of the weights of these rabbits. **(3)**

Given that the median weight of these rabbits is 815 grams,

(b) describe, giving a reason, the skewness of these data. **(2)**

Two more rabbits weighing 776 grams and 896 grams are added to make a group of 12 rabbits.

(c) State, giving a reason, how the inclusion of these two rabbits would affect the mean. **(2)**

(d) By considering the change in $\sum (x - \bar{x})^2$, state what effect the inclusion of these two rabbits would have on the standard deviation. **(2)**

6(a) $\bar{x} = \frac{\sum x}{n}$

$= \frac{8360}{10}$

$\therefore \bar{x} = 836\text{g}$

$\sigma^2 = \frac{\sum (x-\bar{x})^2}{n}$

$= \frac{63840}{10}$

$= 6384$

$\sigma = \sqrt{6384}$

$= 4\sqrt{399}$

$\therefore \sigma \approx 79.9\text{ kg}$ (3 SF)

(b) mean > median $\Leftrightarrow$ positive skewness.

(c) new mean $= \frac{8360 + 776 + 896}{10 + 2}$

$= 836\text{ kg}$

$\therefore$ no change.

Question 6 continued

(d) $S_{xx} = \sum (x - \bar{x})^2$

$S_{xx} = 63840$

$\sum x^2 + \frac{(\sum x)^2}{n} = 63840$

$\sum x^2 + \frac{8360^2}{10} = 63840$

$\sum x^2 = 7052800$

new $\sum x^2 = 7052800 + 776^2 + 896^2$

$= 8457792$

new $\sigma = \sqrt{\frac{\text{new } \sum x^2}{12} - \bar{x}^2}$

$= \sqrt{\frac{8457792}{12} - (836)^2}$

$= \sqrt{5926}$

$= 4\sqrt{370}$

$= 76.9415...$

$\therefore$ new $\sigma \approx 76.9$ kg (3SF)

$\therefore$ standard deviation decreases.

Q6

(Total 9 marks)

TOTAL FOR PAPER: 75 MARKS

END