S1 JAN 14 (*INT)

1. A price comparison website publishes data on the cost per month, £c, and the level of satisfaction, s, of a random sample of six internet service providers. A low value of s corresponds to a low level of satisfaction. The data are given in the table below.

Internet service provider	A	B	C	D	E	F
c	20	15	12	30	9	25
s	5	3	4	2	3	4

(You may use $\sum c = 111$, $\sum c^2 = 2375$, $\sum s = 21$, $\sum s^2 = 79$, $\sum cs = 380$, $S_{cc} = 321.5$)

(a) Calculate the value of S_{cs} and the value of S_{ss} (3)

(b) Calculate the product moment correlation coefficient for these data. (2)

Brad is not satisfied with his current internet service and decides to change his provider. He decides to pay a lot more for his new internet service.

(c) On the basis of your calculation in part (b), comment on Brad's decision. Give a reason for your answer. (2)

a) $S_{cs} = 380 - 111 \times 21 \div 6 = -8.5$

$S_{ss} = 79 - 21^2 \div 6 = 5.5$

b) $r = \dfrac{-8.5}{\sqrt{5.5 \times 321.5}} = -0.202$

c) There seems to be very little correlation between costs and satisfaction. However, people paying more may expect more. So he may get a better service but is unlikely to be more satisfied

2. A rugby club coach uses club records to take a random sample of 15 players from 1990 and an independent random sample of 15 players from 2010. The body weight of each player was recorded to the nearest kg and the results from 2010 are summarised in the table below.

x	77	82	87	92	97	102	107
Body weight (kg)	75–79	80–84	85–89	90–94	95–99	100–104	105–109
f Number of Players (2010)	1	2	2	4	3	2	1

cf 0 1 3 5 9 12 14 15

(a) Find the estimated values in kg of the summary statistics a, b and c in the table below.

	Estimate in 1990	Estimate in 2010
Mean	83.0	a
Median	82.0	b
Variance	44.0	c

Give your answers to 3 significant figures. (6)

The rugby coach claims that players' body weight increased between 1990 and 2010.

(b) Using the table in part (a), comment on the rugby coach's claim. (2)

a) mean $= \dfrac{\sum fx}{\sum f} = \dfrac{1385}{15} = 92\frac{1}{3}$

b) $\frac{1}{2}n = 7.5$

89.5 Q_2 94.5

5 7.5 9

$\dfrac{Q_2 - 89.5}{5} = \dfrac{2.5}{4} \Rightarrow Q_2 = 92.625$

c) $Var = \dfrac{S_{xx}}{n} = \dfrac{128855 - 1385^2 \div 15}{15}$

$Var(X) = 64.8$

d) the average weight increased by about 10kg but spread has also increased

3. Jean works for an insurance company. She randomly selects 8 people and records the price of their car insurance, £p, and the time, t years, since they passed their driving test. The data is shown in the table below.

t	10	13	17	18	22	24	25	27
p	720	650	430	490	500	390	280	300

(You may use $\bar{t} = 19.5$, $\bar{p} = 470$, $S_{tp} = -6080$, $S_{tt} = 254$, $S_{pp} = 169\,200$)

(a) On the graph below draw a scatter diagram for these data. (2)

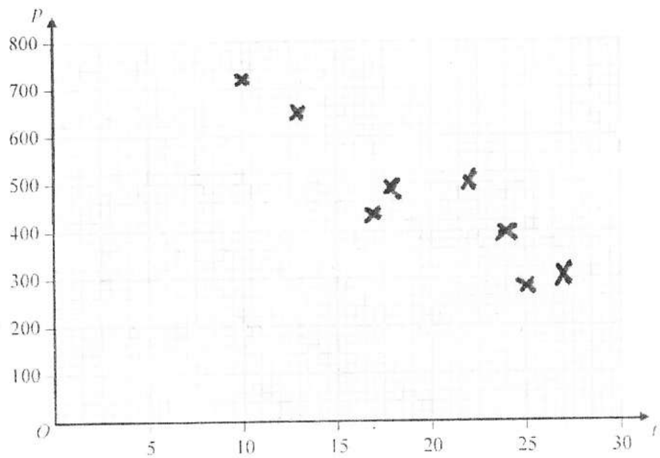
(b) Comment on the relationship between p and t. (1)

(c) Find the equation of the regression line of p on t. (4)

(d) Use your regression equation to estimate the price of car insurance for someone who passed their driving test 20 years ago. (2)

Jack passed his test 39 years ago and decides to use Jean's data to predict the price of his car insurance.

(e) Comment on Jack's decision. Give a reason for your answer. (2)



b) negative correlation.
as time increases, insurance cost decreases.

c) $b = \dfrac{S_{tp}}{S_{tt}} = -23.937..$

$a = \bar{p} - b\bar{t} = 936.7717..$

$p = 936.8 - 23.9t$

d) $t = 20$ $\quad p = £458$

e) unreliable, extrapolation, outside of data range.

4. A discrete random variable X has the probability distribution given in the table below, where a and b are constants.

x	-1	0	1	2	3
$\mathrm{P}(X = x)$	a	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	b

Given $\mathrm{E}(X) = \frac{9}{5}$

(a) (i) find two simultaneous equations for a and b.

(ii) show that $a = \frac{1}{20}$ and find the value of b. (4)

(b) Specify the cumulative distribution function $\mathrm{F}(x)$ for $x = -1, 0, 1, 2$ and 3 (2)

(c) Find $\mathrm{P}(X < 2.5)$. (1)

(d) Find $\mathrm{Var}(3 - 2X)$. (4)

$a + b + 0.6 = 1 \Rightarrow \quad a + b = 0.4$

$E(x) = -a + 0.2 + 0.6 + 3b = 1.8 \Rightarrow -a + 3b = 1 \quad +$

$4b = 1.4$

$b = 0.35$

$\therefore a = 0.05$

b)

x	-1	0	1	2	3
$F(x)$	0.05	0.15	0.35	0.65	1

c) $P(X<2.5) = 0.65$

d) $V(3-2X) = (-2)^2 V(X)$

$V(X) = E(X^2) - E(X)^2$

$V(X) = (0.05+0.2+1.2+3.15) - 1.8^2$

$V(X) = 1.36$

$\therefore V(3-2X) = \underline{5.44}$

5. A group of 100 students are asked if they like folk music, rock music or soul music.

All students who like folk music also like rock music

No students like both rock music and soul music

75 students do not like soul music

12 students who like rock music do not like folk music

30 students like folk music

(a) Draw a Venn diagram to illustrate this information. **(4)**

(b) State two of these types of music that are mutually exclusive. **(1)**

Find the probability that a randomly chosen student

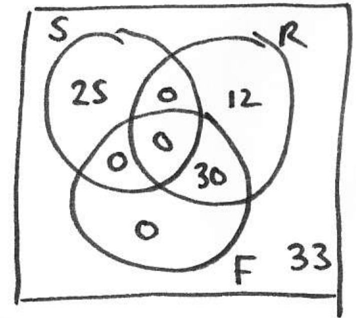
(c) does not like folk music, rock music or soul music, **(1)**

(d) likes rock music, **(1)**

(e) likes folk music or soul music. **(1)**

Given that a randomly chosen student likes rock music,

(f) find the probability that he or she also likes folk music. **(2)**



b) Soul and Rock

c) $\frac{33}{100}$

d) $\frac{42}{100}$

e) $\frac{55}{100}$

f) $\frac{30}{42}$

6. A manufacturer has a machine that fills bags with flour such that the weight of flour in a bag is normally distributed. A label states that each bag should contain 1 kg of flour.

(a) The machine is set so that the weight of flour in a bag has mean 1.04 kg and standard deviation 0.17 kg. Find the proportion of bags that weigh less than the stated weight of 1 kg. **(3)**

The manufacturer wants to reduce the number of bags which contain less than the stated weight of 1 kg. At first she decides to adjust the mean but not the standard deviation so that only 5% of the bags filled are below the stated weight of 1 kg.

(b) Find the adjusted mean. **(3)**

The manufacturer finds that a lot of the bags are overflowing with flour when the mean is adjusted, so decides to adjust the standard deviation instead to make the machine more accurate. The machine is set back to a mean of 1.04 kg. The manufacturer wants 1% of bags to be under 1 kg.

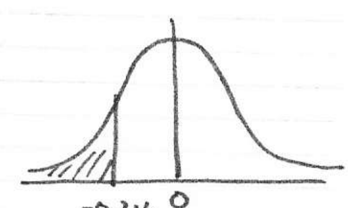
(c) Find the adjusted standard deviation. Give your answer to 3 significant figures. **(3)**

$X \sim N(1.04, 0.17^2)$

$P(X<1)$

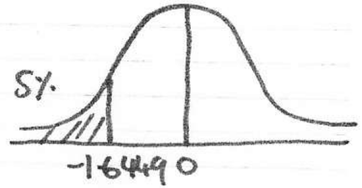
$= P\left(Z < \frac{1-1.04}{0.17}\right)$

$= P(Z<-0.24)$



$= \Phi(-0.24) = 1 - \Phi(0.24) = \underline{0.4052}$

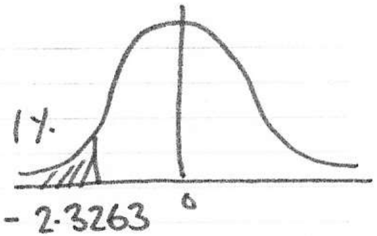
$P(X<1) = 0.05$



$P\left(Z < \frac{1-\mu}{0.17}\right) = 5\%$

$\frac{1-\mu}{0.17} = -1.6449$ $\Rightarrow 1-\mu = -0.2796$ $\therefore \mu = 1.28$

c) $P(X<1) = 1\%$



$P\left(Z < \frac{1-1.04}{\sigma}\right) = 1\%$

$-\frac{0.04}{\sigma} = -2.3263$ $\Rightarrow \sigma = 0.0172$

7. In a large college, $\frac{3}{5}$ of the students are male, $\frac{3}{10}$ of the students are left handed and $\frac{1}{5}$ of the male students are left handed.

A student is chosen at random.

(a) Given that the student is left handed, find the probability that the student is male. **(2)**

(b) Given that the student is female, find the probability that she is left handed. **(3)**

(c) Find the probability that the randomly chosen student is male and right handed. **(2)**

Two students are chosen at random.

(d) Find the probability that one student is left handed and one is right handed. **(2)**

	Left	Right	total
Male	1/5 of 0.6 0.12	0.48	0.6
female	0.18	0.22	0.4
total	0.3	0.7	1

a) $\frac{0.12}{0.3} = \frac{2}{5}$

b) $\frac{0.18}{0.4} = \frac{9}{20}$

c) 0.48

d) LR or RL $= 0.3 \times 0.7 + 0.7 \times 0.3$

$= 0.42$

8. A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

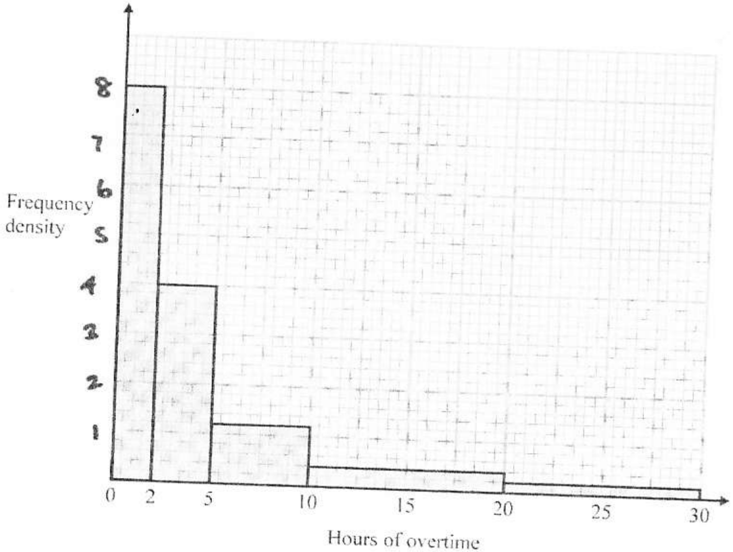


Figure 1

(a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month. **(4)**

(b) Estimate the median number of hours of overtime claimed by these 40 staff in the month. **(2)**

(c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month. **(2)**

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

(d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff. **(2)**

Area $\propto$ freq

$(2\times8)+(3+4)+(5\times1.2)+(10\times0.4)+(10\times0.2)$

$\Rightarrow 40 = 40k \quad \therefore k=1 \quad$ Area = freq

a) $16+12+6 = 34$

b)

	f	x	cf
			0
0-2	16	1	16
2-5	12	3.5	28
5-10	6	7.5	34
10-20	4	15	38
20-30	2	25	40

$\frac{1}{2}n = 20$

2 Q_2 5

16 20 28

$\frac{Q_2 - 2}{3} = \frac{4}{12} \qquad \underline{Q_2 = 3}$

c) $\Sigma fx = 213 \qquad \Sigma fx^2 = 2650.5$

mean $= \frac{213}{40} = \underline{5.3}$

d) Median as the data is heavily skewed.