1) i) To estimate the mean and standard deviation for grouped data, we assume that the data in each group is at the midpoint in each group

Area (x)	Frequency	Midpoint(x)	$f \times x$	$f \times x^2$
$0 < x \leq 3$	3	1.5	4.5	6.75
$3 < x \leq 5$	8	4	32	128
$5 < x \leq 7$	13	6	78	468
$7 < x \leq 10$	14	8.5	119	1011.5
$10 < x \leq 20$	6	15	90	1350
	$\Sigma f = 44$		$\Sigma fx = 323.5$	$\Sigma fx^2 = 2964.25$

Estimate for mean $= \dfrac{\Sigma fx}{n}$ where $n = \Sigma f$

$$= \frac{323.5}{44} = 7.352272727$$

$$= 7.35 \quad \text{to 3 sig fig}$$

Estimate for s.d.

$$= \sqrt{\frac{\Sigma fx^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{2964.25 - 44 \times 7.352272727^2}{43}}$$

$$= 3.6909 = 3.69 \quad \text{to 3 sig fig}$$

Note (It is a good idea to use several decimal places in your $\bar{x}$ in this calculation because rounding errors can take you outside the range of answers that gain full marks)

---

1 ii) At top end outliers are more than $2 \times$ s.d above $\bar{x}$

$7.35 + 2 \times 3.69 = 14.73$

Since for top group $10 < x \leq 20$ there could be outliers above 14.73

---

2) i) W is event person speaks Welsh
C is event person is a child

$P(W) = 0.20$, $P(C) = 0.17$, $P(W \cap C) = 0.06$

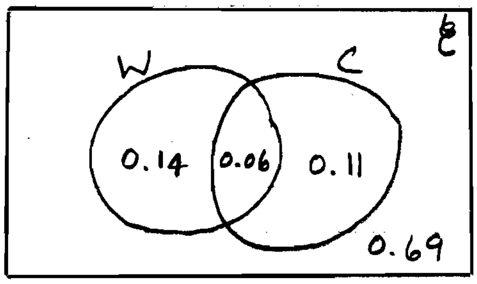
Events are independent if and only if $P(W) \times P(C) = P(W \cap C)$

But $P(W) \times P(C) = 0.20 \times 0.17 = 0.034 \neq 0.06$

$\therefore$ events W and C are <u>not</u> independent

---

2 ii)



First we put in 0.06 the prob of the intersection which we were given
Then since $P(W) = 0.20$ we can work out $0.20 - 0.06 = 0.14$ to give the prob in the other part of W. In the same way we can work out the prob for the other part of C. $0.17 - 0.16 = 0.11$
Finally, all the probs have to add up to 1. So the prob outside the circles $= 1 - (0.14 + 0.06 + 0.11) = 0.69$

---

2 iii)

$$P(W \mid C) = \frac{P(W \cap C)}{P(C)} = \frac{0.06}{0.17} = 0.353 \text{ to 3 d.p.}$$

Remember this as: The probability of one event given that another has occurred is given by the probability of their intersection divided by the probability of the one that has occurred.

---

2 iv)

$P(W \mid C) = 0.353$ Given the person is a child, the prob that they speak Welsh is 0.353

$P(W \mid C') = 0.169$ Given the person is not a child ($\therefore$ an adult) the prob they speak Welsh is 0.169

Conclusion: A higher proportion of children speak Welsh than adults.

<u>or</u> The proportion of children that speak Welsh is about twice the proportion of adults.

---

3) i)

A)

r	0	1	2	3
$P(X=r)$	0·5	0·35	p	q

$\sum_{r=0}^{3} P(X=r) = 1$ i.e. Probabilites add up to 1

$\therefore \quad 0.5 + 0.35 + p + q = 1$

$0.85 + p + q = 1$

$p+q = 1 - 0.85$

$p+q = 0.15$

B)

$E(x) = 0.5 \times 0 + 0.35 \times 1 + p \times 2 + q \times 3$

$\therefore 0.67 = 0.35 + 2p + 3q$

$\therefore 0.67 - 0.35 = 2p + 3q$

$2p + 3q = 0.32$

c)

$p + q = 0.15$ ①

$2p + 3q = 0.32$ ②

①×2 $2p + 2q = 0.30$ ③

②-③ $q = 0.02 \Rightarrow p = 0.13$

3ii) $Var(x) = E(x^2) - (E(x))^2$

$E(x^2) = 0.5 \times 0^2 + 0.35 \times 1^2 + 0.13 \times 2^2 + 0.02 \times 3^2 = 1.05$

$E(x) = 0.5 \times 0 + 0.35 \times 1 + 0.13 \times 2 + 0.02 \times 3 = 0.67$

$Var(x) = 1.05 - 0.67^2$

$Var(x) = 0.6011$

4) i) $X \sim B(8, 0.05)$

A) $P(X=0) = 0.6634$ (from tables) OR $P(X=0) = 0.95^8 = 0.6634$

B) $P(X>1) = 1 - P(X \leq 1)$

$= 1 - 0.9428$ (from tables)

$= 0.0572$

4ii) Expected days with more than one off sick $= 250 \times 0.0572$

$= 14.3$

5) i) $X \sim B(15, 0.35)$

$H_0 : p = 0.35$
$H_1 : p > 0.35$

where p is prob that student names all items

H_1 is $p > 0.35$ because improvement is anticipated in this case.

ii) Expected value $np = 15 \times 0.35 = 5.25$

Actual value $X = 8$

Consider $P(X \geq 8)$

$P(X \geq 8) = 1 - P(X \leq 7)$

$= 1 - 0.8868 = 0.1132$

Since $0.1132 > 5\%$ there is not sufficient evidence to reject H_0

Conclusion: There is not sufficient evidence to suggest that listening to the music increases the probability of remembering all the items

Section B

6) i) A) P(Both born in rest of UK) $= 0.20 \times 0.20 = 0.04$

B) P(At least one born in England) $= 1 -$ P(Neither born in England)

$= 1 - 0.21 \times 0.21$

$= 0.9559$

c) P(Neither born overseas) $= 0.99 \times 0.99 = 0.9801$

6ii) P(Both born in rest of UK | neither born overseas)

$$= \frac{P(\text{Both born in rest of UK} \cap \text{neither born overseas})}{P(\text{neither born overseas})}$$

(Note: the top probability is the same as P(Both born in rest of UK) because people born in rest of UK are also not born overseas.)

$$= \frac{0.04}{0.9801} = 0.0408$$

6iii) A) P(At least one not born in England) $= 1 -$ P(All 5 born in England)

$= 1 - 0.79^5 = 0.6923$

B) For n people P(At least one not born in England) $= 1 - 0.79^n$

We require

$1 - 0.79^n > 0.9$ (90%)

$1 - 0.9 > 0.79^n$

$0.1 > 0.79^n$

By trial and error $0.79^9 = 0.1199$ so select $n = 10$ people

$0.79^{10} = 0.0946$

7) i) Positive skew

ii) People under 25 given by $33000 \times 20 + 58000 \times 5$

$= 950,000$

iii)

A) Cumulative freq up to 40 = 1810 thousands

Extra between 40 and 50 $= 34 \times 10 = 340$ thousands

$\therefore$ a = cumulative freq up to 50 = 1810 + 340 thousands

= 2150 thousands

$\therefore a = 2150$

B)

$\frac{2770}{2} = 1385$ 1385 occurs in 30 to 40 interval

$1385 - 1240$ (beginning of interval)
$= 145$

$1810 - 1240$ (width of interval)
$= 570$

Estimate of median = age at beginning of interval + appropriate fraction of interval

$= 30 + \frac{145}{570} \times 10$

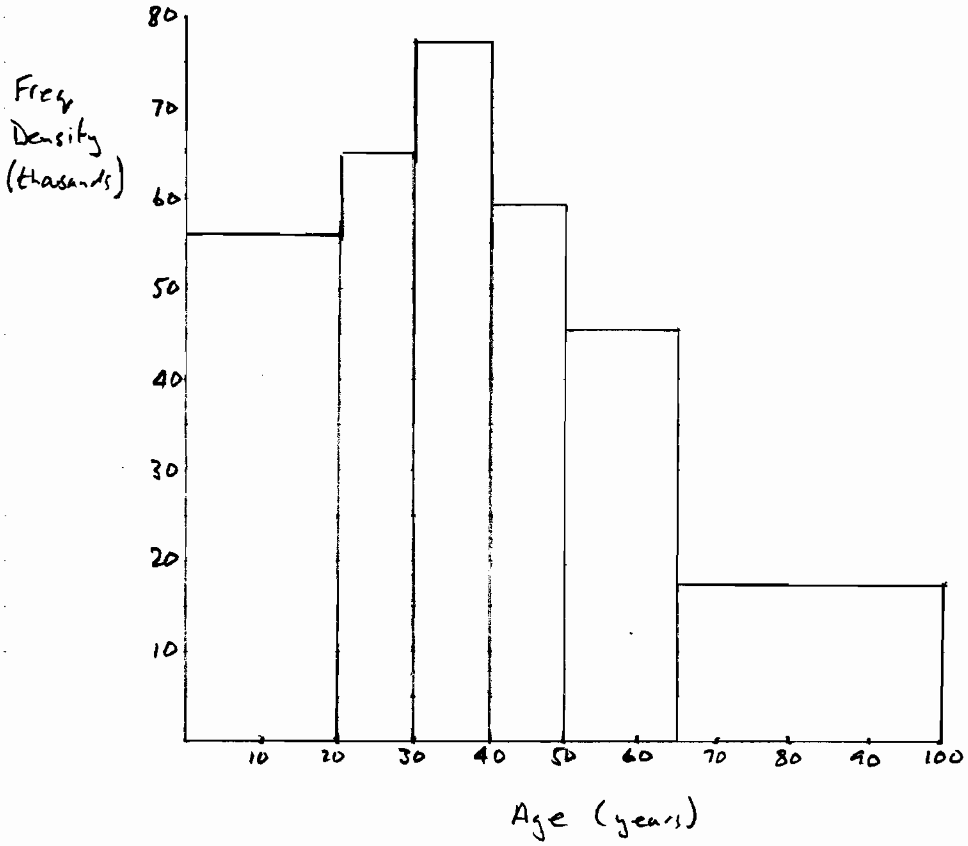
$= 32.54$ years

$= 32.5$ to 3 sig fig

(Note: Do not round to a whole number. Expected values do not have to be whole numbers.)

7 cont) iv)

Age	$0 \le x < 20$	$20 \le x < 30$	$30 \le x < 40$	$40 \le x < 50$	$50 \le x < 65$	$65 \le x < 100$
Freq (th)	1120	650	770	590	680	610
Freq Density	56	65	77	59	45.33	17.43



7v)

Modal group for Inner London is 20-30 but 30-40 for Outer London.
Outer London has a greater proportion of aged 65+ residents than Inner London.
Any reasonable comments would gain marks!

7vi)

Mean 38.5, Median 35.7, midrange 50, s.d 23.7, IQR 34.4

Recalculating with assumed max age rising to 105 years

Mean increases (midpoint of top group would be higher)

Median unchanged (still middle person's age)

Midrange increases (since top of range increases)

s.d increases (data more spread)
IQR unchanged (since it only considers middle 50%)