1. George throws a ball at a target 15 times.
Each time George throws the ball, the probability of the ball hitting the target is 0.48

The random variable X represents the number of times George hits the target in 15 throws.

(a) Find

(i) $P(X = 3)$

(ii) $P(X \geqslant 5)$

(3)

George now throws the ball at the target 250 times.

(b) Use a normal approximation to calculate the probability that he will hit the target more than 110 times.

(3)

X follows a Binomial Distribution $\sim B(15, 0.48)$

a) i) $P(X=3) = 0.48^3 \times 0.52^{12} \times {}_{15}C_3$

$= 0.1106 \times 0.00003908 \times 455$

$= \underline{0.0197}$

ii) $P(X \geqslant 5) = 1 - [P(X=1) + P(X=2) + P(X=3) + P(X=4)]$

(cumulative probability)

$= 1 - 0.0799$

$= \underline{0.920}$

b) Approximate $B(250, 0.48)$ by a normal distribution (this is OK because the number of trials is large and the probability is close to 0.5

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

Question 1 continued

To use the normal distribution

take mean $\mu = np = 250(0.48)$
$= 120$

& standard dev^n $\sigma = \sqrt{npq} = \sqrt{250(0.48)(0.52)}$
$= \sqrt{62.4} = 7.90$

So $X \sim N(120, 7.90^2)$

So according to the normal dist^n

$P(X > 110) = 1 - P(X \leq 110)$

$= \ldots$ except that we need to think discrete intervals,* so we use $P(X \leq 110.5)$

which gives $1 - 0.1146$

$= \underline{0.8854}$

*this is the 'continuity correction'

(Total for Question 1 is 6 marks)

P72130A0320

Turn over

2. A manufacturer uses a machine to make metal rods.

The length of a metal rod, L cm, is normally distributed with

- a mean of 8 cm
- a standard deviation of x cm

Given that the proportion of metal rods less than 7.902 cm in length is 2.5%

(a) show that $x = 0.05$ to 2 decimal places.

(2)

(b) Calculate the proportion of metal rods that are between 7.94 cm and 8.09 cm in length.

(1)

The **cost** of producing a single metal rod is 20p

A metal rod

- where $L < 7.94$ is **sold** for scrap for 5p
- where $7.94 \leqslant L \leqslant 8.09$ is **sold** for 50p
- where $L > 8.09$ is shortened for an extra **cost** of 10p and then **sold** for 50p

(c) Calculate the expected profit per 500 of the metal rods.
Give your answer to the nearest pound.

(5)

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of 0.015 of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

(d) Explain whether the manufacturer is likely to achieve its aim.

(4)

$L \sim N(8, x^2)$

a) we know $P(L < 7.902) = 0.025$

we need to standardise the normal dist^n

So write $Z = \dfrac{L - \mu}{\sigma} = \dfrac{L-8}{x}$ 1

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

Then we know for Z:

$P(Z < t) = 0.025$

gives an inverse distribution value

of $t = -1.960$

So substituting back into ①:

$\frac{L-8}{x} = -1.960$ when $L = 7.902$

so $\frac{7.908-8}{x} = -1.960$

$x = 0.047 \approx 0.05$ as required.

b) $P(L \leq 7.94) = 0.1151$

$P(L \leq 8.09) = 0.9641$

So $P(7.94 \leq L \leq 8.09) = 0.9641 - 0.1151$

$= \underline{0.849}$.

c)

	Cost	Sold	Profit	P(L=)
$L < 7.94$	20p	5p	−15p	0.115
$7.94 \leq L \leq 8.09$	20p	50p	30p	0.849
$8.09 < L$	30p	50p	20p	0.036

So expected profit on 500 rods

$= \Sigma$ profit $\times$ probability

$= 500(-15 \times 0.115 + 30 \times 0.849 + 20 \times 0.036)$

$= 500(-1.725 + 25.47 + 0.72)$

$= 500 \times 24.465$

$= 12232.5p$

$= £122$ to the nearest £.

d) The distribution here is $X \sim B(200, 0.015)$

and the manufacturer wants

$P(X < 6) < 95\% = 0.95$

this is $P(X \leq 5) = 0.9176$ by calculation.

So $P(X \leq 5)$ is only $92°$

not $95°$ as the manufacturer

wanted.

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

3. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution B(14, 0.27) might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

(1)

a) Trace.

b) i) mean $\mu = \frac{\Sigma r}{n} = \frac{174.9}{31} = \underline{5.64\text{mm}}$

ii) standard deviation $= \sqrt{\text{variance}}$

variance $= \frac{\Sigma r^2}{n} - \mu^2$

$= \frac{3523.283}{31} - 5.6419^2$

$= 81.823$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

so standard deviation $= \sqrt{81.823}$

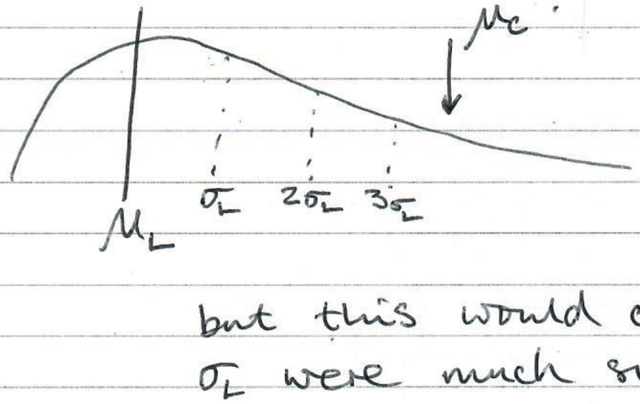
$= \underline{9.0456}$ mm

c) (Dion must be a real plonker...)

Mean rainfall in Camborne in Aug = 5.64mm
— — — — Leuchars ... 1.72mm

The mean in Leuchars is far below the mean in Camborne, so there is no reason (ie no evidence) to support Dion's belief.

(It's curious that $\mu = 5.64$ but $\sigma = 9.05$ - i.e much larger. Yet if we were looking for a '3x SD' argument we clearly couldn't expect the mean at Leuchars to be $< \mu - 3\sigma$ since this is very negative. Perhaps the way to frame this would be to look at the rainfall at Leuchars:



but this would only work if σ_L were much smaller than σ_C.

d) Note this is for the <u>whole year</u>. There is no reason to think it's true for August - ie one month - when there (1) may be less, or more, rain. (2) $B(14, 0.27)$ doesn't particularly match $N(\mu, \sigma^2)$, (3) it's questionable whether the rainfall is a normal or a binomial dist^n anyway; (4) this is presumably a <u>continous</u> 14-day event.

(Total for Question 3 is 7 marks)

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1)

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)
You should state the probability of rejection in each tail, which should be less than 0.025

(3)

(c) Find the actual level of significance of the test based on your critical region from part (b)

(1)

The manager observes that 15 of the 50 customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief.

(1)

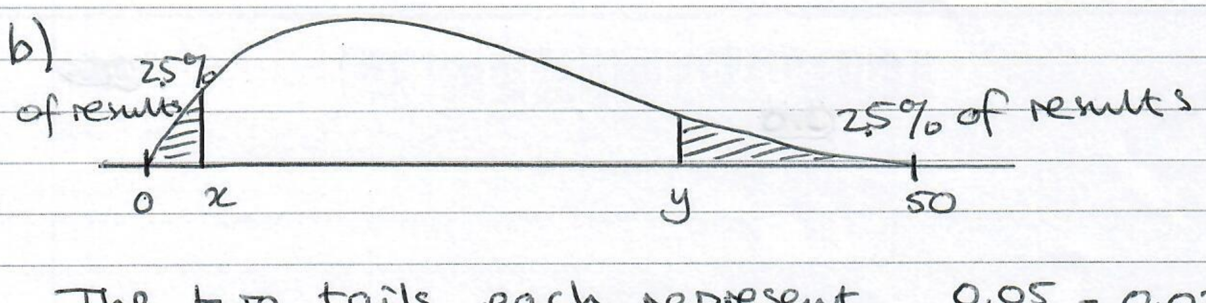
$X \sim B(50, 0.1)$

a) H_0: no change : $p = 0.1$

H_1: a significant change. $p \neq 0.1$

use a two-tailed test because we don't know in which direction it might have changed.

b)



The two tails each represent $\frac{0.05}{2} = 0.025$

probability of the value of X

Question 4 continued

At the left end,

Binom.inv (0.025, 50, 0.1) = 1

and in fact Binom (1, 50, 0.1) = 0.34
(cumul)

so the 2.5% region is < 1 - i.e. $x=0$.
so the critical value x is 0.
Binom (0, 50. 0.1) = 0.0052
At the right end,

Binom.inv (0.975, 50, 0.1) = 9
Binom (8, 50, 0.1) = 0.9421 **
and Binom (9, 50. 0.1) = 0.9755 *
(cumul)

Binom (10, 50, 0.1) = 0.9906

* tells us P(10 or more) = 0.0245 (2)
** — — — P(9 or more) = 0.0579

So the critical value y is 10.

Probabilities:

$P(X=0) = 0.0052$

$P(X \geq 10) = 0.245$.

I'm not convinced these are "probabilities of rejection" – rather I think they are 'the critical regions which would make us believe H_1'

c) the actual significance level
= 0.0052 + 0.0245 = 0.0297 (I don't understand the examiner's comment that

we'd expect this to be near 0.025 ... surely it should 'ideally' be near 0.05.

d) A score $x = 15$ lies well within the critical region, so we reject H_0 and - -

accept the manager's claim that the proportion of customers who arrive late has changed.' (note the form of words)

5. A company has 1825 employees.
The employees are classified as professional, skilled or elementary.

The following table shows

- the number of employees in each classification
- the two areas, A or B, where the employees live

	A	B	Total
Professional	740	380	1120
Skilled	275	90	365
Elementary	260	80	340
Total	1275	550	1825.

An employee is chosen at random.

Find the probability that this employee

(a) is skilled,

(1)

(b) lives in area B and is not a professional.

(1)

Some classifications of employees are more likely to work from home.

- 65% of professional employees in both area A and area B work from home 1
- 40% of skilled employees in both area A and area B work from home 2
- 5% of elementary employees in both area A and area B work from home 3
- Event F is that the employee is a professional 4
- Event H is that the employee works from home 5
- Event R is that the employee is from area A 6

(c) Using this information, complete the Venn diagram on the opposite page.

(4)

(d) Find $P(R' \cap F)$

(1)

(e) Find $P([H \cup R]')$

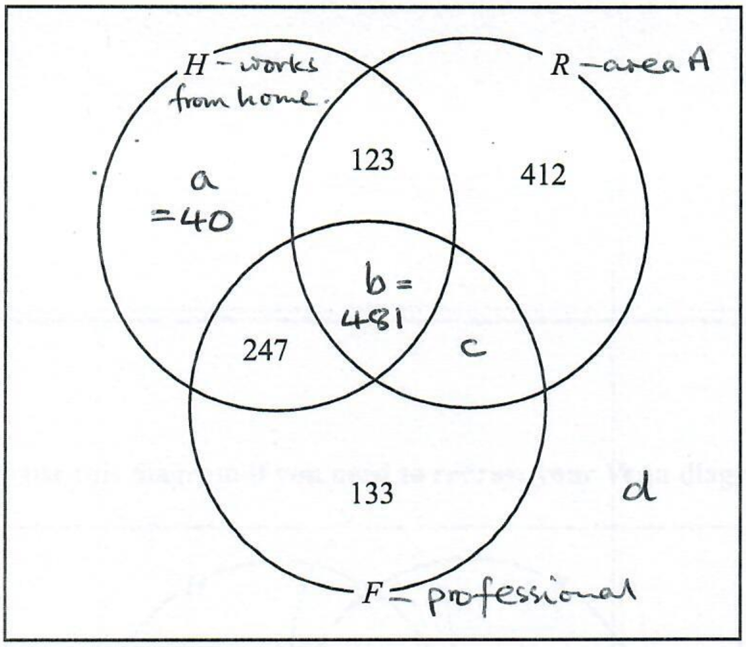
(1)

(f) Find $P(F \mid H)$

(2)

a) 365/1865 = 20% (0.2)
b) 170/1865 = 0.9115

Question 5 continued



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

Using the equations in the question:

① $.65(1120) = 728 = 247 + b$ $\quad \underline{b = 481}$

② $.40(365) = 146$

③ $.05(340) = 17$

$146 + 17 = 163 = a + 123$

$\underline{\underline{a = 40}}$.

And we know $b + C = 740$

so $C = 740 - 481 = \underline{\underline{259}}$.

Lastly $d = 1825 - (40 + 123 + 412 + 247 + 481 + 259 + 133)$

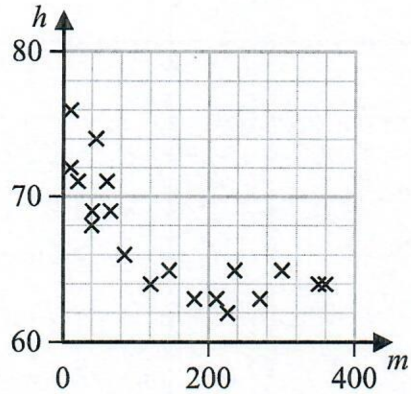
$= 1825 - 1695$

$\underline{\underline{d = 130}}$

6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram.



(a) Interpret the nature of the relationship between h and m

(1)

Anna codes the data using the formulae

$$x = \log_{10} m$$
$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

(b) Test whether or not there is significant evidence of a negative correlation between x and y
You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

(c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

Question 6 continued

a) There is a negative correlation between h and m.

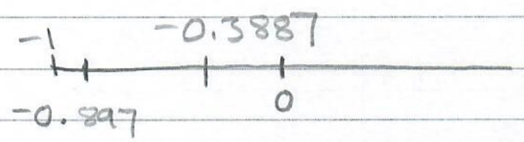
(It's not over-clear, and it might look more convincing on different scales - but this isn't required.)

b) Hypotheses: Let C be the correlation between x and y.

H_0 : $C = 0$

H_1 : $C < 0$ (or actually, is significantly enough below 0 to make us believe the negative correlation).

To test for significant correlation on 19 samples, use the correlation coefficient table for 0.05 (5%) significance with one tail (negative) – this gives 0.3887



The critical region is $[-1, -0.3887]$

so the measured value -0.897 is within this region

So there is sufficient reason to reject H_0 and conclude that there is a negative correlation between x and y.

Question 6 continued

c) $y = -0.05x + 1.92$

so $\log h = -0.05 \log m + 1.92$

(base 10 implied)

so $h = m^{-0.05} \times 10^{1.92}$

$h = 10^{1.92} m^{-0.05}$

$h = 83.18 m^{-0.05}$

which is in the required form
with $a = 83.18$
$k = -0.05$.

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