Edexcel S1 Jan 2002

5. The duration of the pregnancy of a certain breed of cow is normally distributed with mean μ days and standard deviation σ days. Only 2.5% of all pregnancies are shorter than 235 days and 15% are longer than 286 days.

~~(a)~~ Show that $\mu - 235 = 1.96\sigma$. **(2)**

~~(b)~~ Obtain a second equation in μ and σ. **(3)**

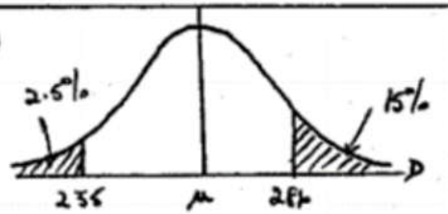
~~(c)~~ Find the value of μ and the value of σ. **(4)**

(d) Find the values between which the middle 68.3% (90) of pregnancies lie. **(2)**

84.15%

$\dfrac{100-68.3}{2}$

5. (a)



$P(D < 235) = 0.025$

$\therefore \dfrac{235-\mu}{\sigma} = -1.96$ — $\dfrac{235-\mu}{\sigma}$ = z — M1

$* \therefore \mu - 235 = 1.96\sigma *$ — A1 (2)

(b) $P(D > 286) = 0.15$ — $\dfrac{286-\mu}{\sigma}$ = z — M1

$\therefore \dfrac{286-\mu}{\sigma} = 1.0364$ $\quad \therefore 286 - \mu = 1.0364\sigma$ — 1.0364 B1 — A1 (3)

(c) Solving for μ or σ — M1

Substituting for other unknown — M1

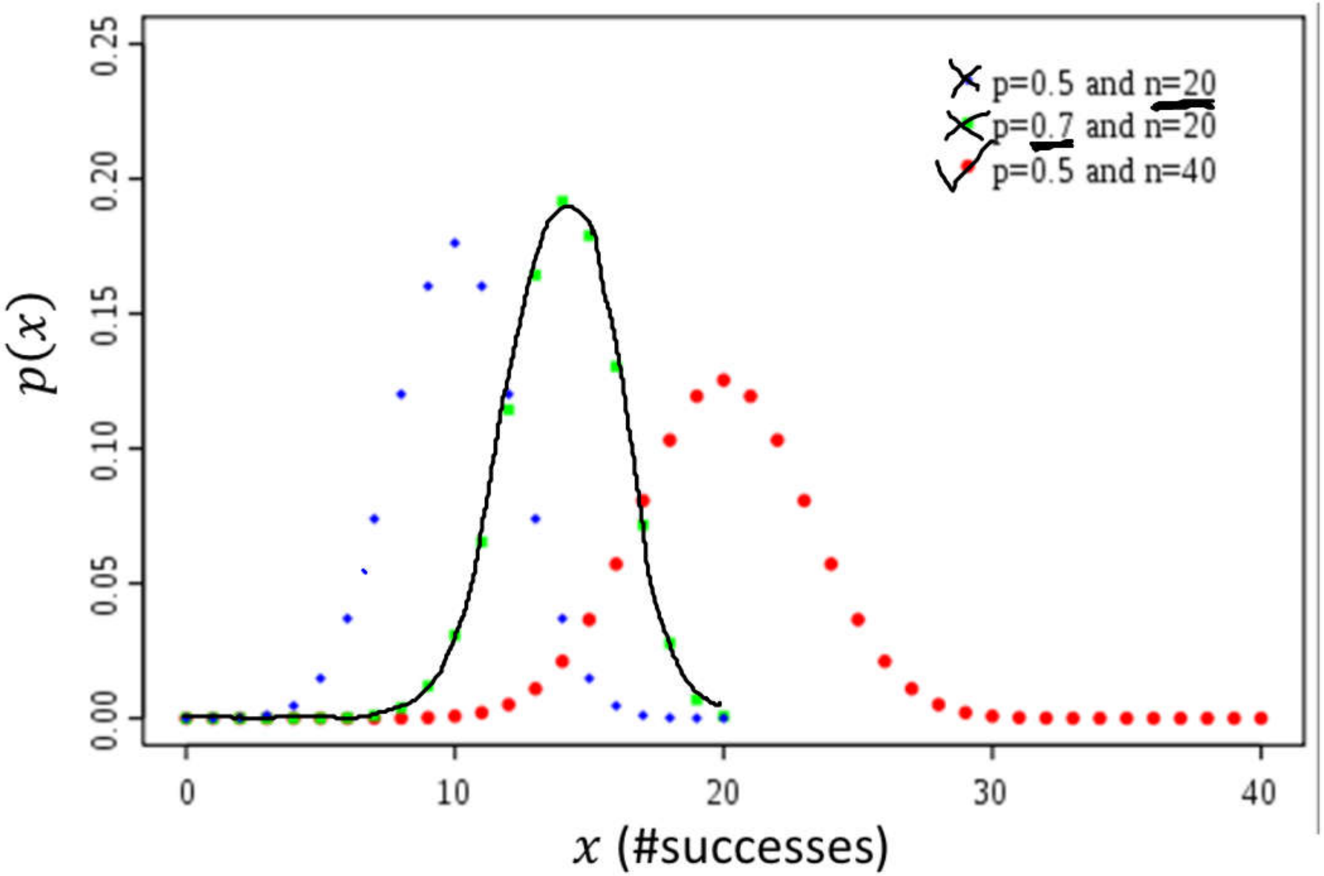
$\mu = 268.3\ldots$ $\sigma = 17.0204\ldots$ — AWRT 268 — A1; AWRT 17 — A1 (4)

(d) $\mu \pm \sigma = 268.36 \pm 17.02$ — μ ± their σ — M1

$= (251, 285)$ — 3sf — A1 (2)

Ex 3E

Approximating a Binomial Distribution



The graph shows the probability function for different Binomial Distributions. Which one resembles another distribution and what distribution does it resemble?

When p is close to 0.5, and n is fairly large, it resembles a normal distribution. The $p = 0.5$ results in the distribution being symmetrical. e.g. For a fair coin toss with 10 throws, we're just as likely to get 1 Head out fo 10 as we are 1 Tail.

If we're going to use a normal distribution to approximate a Binomial distribution, it makes sense that we set the mean and standard deviation of the normal distribution to match that of the original binomial distribution:

$$\boxed{\begin{aligned}\mu &= np \\ \sigma &= \sqrt{np(1-p)}\end{aligned}}$$ memorised.

✏ If n is large and p close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

Quickfire Questions:

$X \sim B(\overset{n}{10}, \overset{p}{0.2}) \rightarrow Y \sim N(2, 1.6)$ $\sigma = \sqrt{1.6}$

$X \sim B(20, 0.5) \rightarrow Y \sim N(10, 5)$ $\sigma = \sqrt{5}$

$X \sim B(6, 0.3) \rightarrow Y \sim N(1.8, 1.26)$ $\sigma = \sqrt{1.26}$

We tend to use the letter Y to represent the normal distribution approximation of the distribution X.

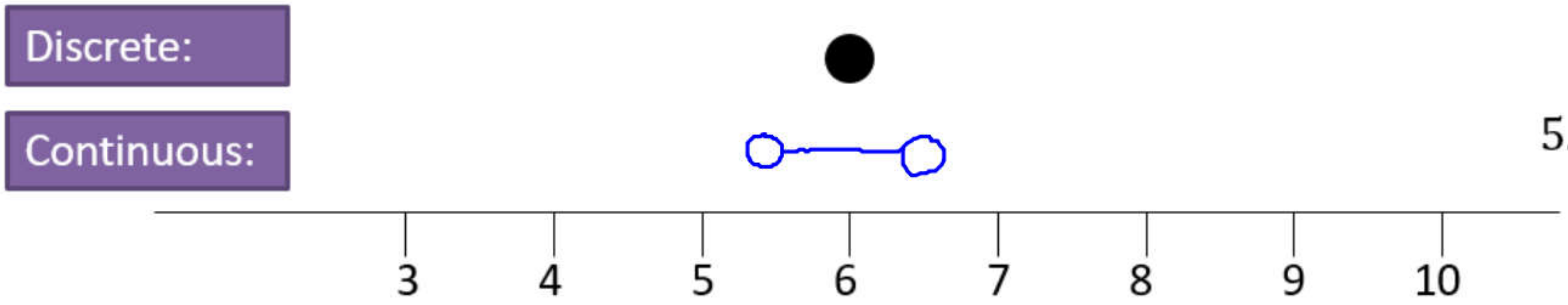
Why use a normal approximation?

- Tables for the binomial distribution only goes up to $n = 50$ and your calculator will reject large values of n.
- The formula for $P(X = x)$ makes use a factorials. Factorials of large numbers cannot be computed efficiently. Type in 65! for example; your calculator will hesitate! Now imagine how many factorials would be required if you wanted to find $P(X \leq 65)$. ☺

Continuity Corrections

One problem is that the outcomes of a binomial distribution (i.e. number of successes) are **discrete** whereas the Normal distribution is **continuous**.
We apply something called a **continuity correction** to approximate a discrete distribution using a continuous one.

The random variable X represents the time to finish a race in hours. We're interested in knowing the probability Alice took 6 hours to the nearest hour. How would you represent this time on a number line given hours is discrete? And what about if hours was now considered to be continuous (as Y)?

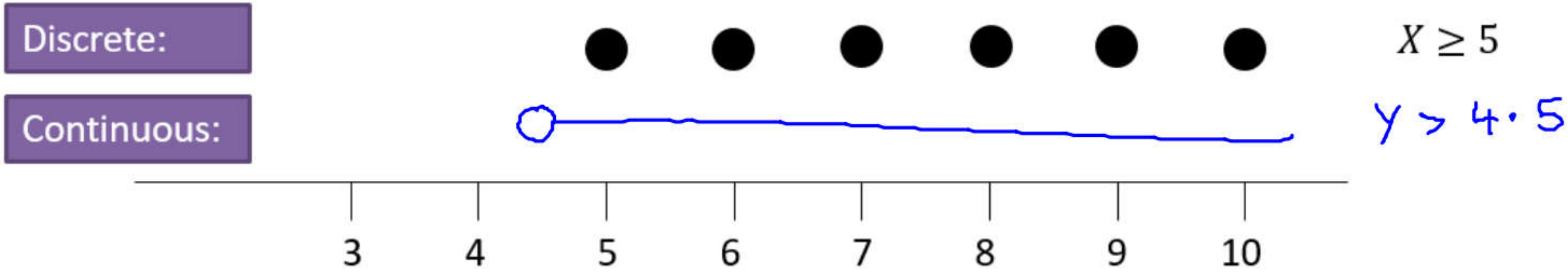


$X = 6$

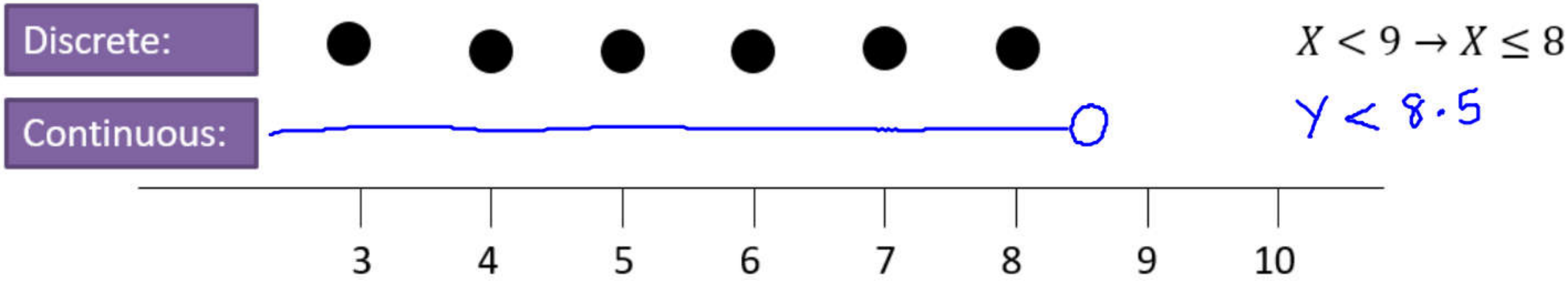
$5.5 < Y < 6.5$

We can't just find $P(Y = 6)$ when Y is continuous, because the probability is effectively 0. But $P(5.5 < Y < 6.5)$ would seem a sensible interval to use because any time between 5.5 and 6.5 would have rounded to 6 hours were it discrete.

If X is a discrete variable, and Y is its continuous equivalent, how would you represent $P(X \geq 5)$ for Y?



How would represent $P(X < 9)$ for Y?



✏ A continuity correction is approximating a discrete range using a continuous one.

1. If > or <, convert to $\geq, \leq$ first.
2. Enlarge the range by 0.5.

Discrete	Continuous
$P(X \le 7)$	$P(Y < 7.5)$
$P(X < 10) = P(X \le 9)$	$P(Y < 9.5)$
$P(X > 9) = P(X \ge 10)$	$P(Y > 9.5)$
$P(1 \le X \le 10)$	$P(0.5 < Y < 10.5)$
$P(3 < X < 6) = P(4 \le x \le 5)$	$P(3.5 < Y < 5.5)$
$P(3 \le X < 6) = P(3 \le x \cancel{8} \le 5)$	$P(2.5 < Y < 5.5)$
$P(3 < X \le 6) = P(4 \le x \le 6)$	$P(3.5 < Y < 6.5)$
$P(X = 3)$	$P(2.5 < Y < 3.5)$

For a particular type of flower bulb~~s~~, 55% will produce yellow flowers. A random sample of 80 bulbs is planted.

(a) Calculate the actual probability that there are exactly 50 yellow flowers.

(b) Use a normal approximation to find a estimate that there are exactly 50 flowers.

(c) Hence determine the percentage error of the normal approximation for 50 flowers.

X is the number of yellow flowers

$$X \sim B(80, 0.55)$$

a) $P(X = 50) = 0.0365$

b) $Y \sim N(np, np(1-p))$

$Y \sim N(44, 19.8)$

$\cancel{P(Y = 50)}$ $\quad P(49.5 < Y < 50.5) = 0.03618$

c) $$\% \text{ error} = \frac{0.03618 - 0.0365}{0.0365} \times 100 = -0.876\ldots$$

$$= -0.88\%$$

% is 0.88% below the true value.

Edexcel S2 Jan 2004 Q3

The discrete random variable X is distributed $B(n, p)$.

(*a*) Write down the value of p that will give the most accurate estimate when approximating the binomial distribution by a normal distribution.
(1)

(*b*) Give a reason to support your value. **(1)**

(*c*) Given that $n = 200$ and $p = 0.48$, find $P(90 \le X < 105)$. **(7)**

a) $p = 0.5$

b) Because when $p = 0.5$, its binomial distribution is symmetrical, like the normal distribution

c) $X \sim B(200, 0.48)$

$P(90 \le X < 105)$

$\boxed{P(89.5 < Y < 104.5)} = \underline{\underline{0.7067}}$

continuity correction.

$Y \sim N(96, 49.92)$

$\mu = 96$

$\sigma = \sqrt{49.92}$

$49.92 \to np(\underline{1-p})$

$1-p \to 0.52$

np

Ex 3F

Q 2, 4, 8

Exam Qs in the booklet.

5. A company sells seeds and claims that 55% of its pea seeds germinate.

(a) Write down a reason why the company should not justify their claim by testing all the pea seeds they produce.

Because if they test them all, they can no longer sell them. (1)

A random selection of the pea seeds is planted in 10 trays with 24 seeds in each tray.

(b) Assuming that the company's claim is correct, calculate the probability that in at least half of the trays 15 or more of the seeds germinate.

Trays ? Seeds in each tray? (3)

(c) Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(1)

A random sample of 240 pea seeds was planted and 150 of these seeds germinated.

n

(d) Assuming that the company's claim is correct, use a normal approximation to find the probability that at least 150 pea seeds germinate.

(3)

(e) Using your answer to part (d), comment on whether or not the proportion of the company's pea seeds that germinate is different from the company's claim of 55%

(1)

b) X is the number of seeds to germinate $X \sim B(24, 0.55)$

$P(X \geq 15) = 1 - P(X \leq 14)$

$= 1 - 0.7009$

$= 0.2991$

T is the number of trays with 15 or more pea seeds germ.

$T \sim B(10, 0.2991)$

$P(T \geq 5) = 1 - P(T \leq 4)$

$= 1 - 0.8513$

$= 0.1487$

c) p is close to 0.5

n is large

d) $X \sim B(240, 0.55)$

$Y \sim N(132, 59.4)$

$\cancel{P(Y > 150)}$

$P(Y > 149.5) = 0.0116 = 1.16\%$

e) Because the probability that the number of seeds was at 150 was very small, the company's claim is not supported

5 (a)	The seeds would be destroyed in the process so they would have none to sell	B1	2.4
		(1)	
(b)	[S = no. of seeds out of 24 that germinate, $S \sim B(24, 0.55)$]		
	T = no. of trays with at least 15 germinating. $T \sim B(10, p)$	M1	3.3
	$p = P(S \geqslant 15) = 0.299126...$	A1	1.1b
	So $P(T \geqslant 5) = 0.1487…$ awrt **<u>0.149</u>**	A1	1.1b
		(3)	
(c)	n is large and p close to 0.5	B1	1.2
		(1)	
(d)	$X \sim N(132, 59.4)$	B1	3.4
	$P(X \geqslant 149.5) = P\left(Z \geqslant \frac{149.5 - 132}{\sqrt{59.4}}\right)$	M1	1.1b
	$= 0.01158…$ awrt **<u>0.0116</u>**	A1cso	1.1b
		(3)	
(e)	e.g The probability is very small therefore there is evidence that the company's claim is incorrect.	B1	2.2b
		(1)	
			(9 marks)

5. A fast food company has a scratchcard competition. It has ordered scratchcards for the competition and requested that 45% of the scratchcards be winning scratchcards.

A random sample of 20 of the scratchcards is collected from each of 8 of the fast food company's stores.

(a) Assuming that 45% of the scratchcards are winning scratchcards, calculate the probability that in at least 2 of the 8 stores, 12 or more of the scratchcards are winning scratchcards.

(5)

(b) Write down 2 conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(1)

A random sample of 300 of the scratchcards is taken. Assuming that 45% of all the scratchcards are winning scratchcards,

(c) use a normal approximation to find the probability that at most 122 of these 300 scratchcards are winning scratchcards.

(4)

Given that 122 of the 300 scratchcards are winning scratchcards,

(d) comment on whether or not there is evidence at the 5% significance level that the proportion of the company's scratchcards that are winning scratchcards is different from 45%

(1)

Question	Scheme	Marks	AOs
5(a)	W = number of scratch cards out of 20 that win, $W \sim \mathrm{B}(20, 0.45)$	B1	3.3
	S=number of stores with at least 12 winning cards $S \sim \mathrm{B}(8, p)$	M1	3.1b
	$p = \mathrm{P}(W \geq 12) = 0.130765$	A1	3.4
	$1 - [\mathrm{P}(S = 1) + \mathrm{P}(S = 0)]$	M1	3.4
	So $\mathrm{P}(S \geq 2) = 0.2818 \ldots$	A1	1.1b
		(5)	
(b)	Number of trials is large and probability of success is close to 0.5	B1	1.2
		(1)	
(c)	$X \sim \mathrm{N}(135, 74.25)$	B1, B1	1.1b,1.1b
	$\mathrm{P}(X < 122.5) = \mathrm{P}\left(Z < \frac{122.5-135}{\sqrt{74.25}}\right)$	M1	3.4
	$= 0.0734 \ldots$	A1	1.1b
		(4)	
(d)	The probability is greater than 0.025 therefore there is insufficient evidence at the 5% significance level to suggest that the proportion is different from 45%	B1	2.2b
		(1)	
			(11 marks)

Notes:
(a) B1 may be implied by subsequent working 1st M1: for selection of appropriate model for S 1st A1: for a correct values of the parameter p 2nd A1: for awrt 0.282
(b) B1: both correct conditions Accept n is large, $np > 5$ and $n(1-p) > 5$
(c) B1: for correct mean B1: for correct variance M1: for continuity correction A1 awrt 0.0734
(d) B1: for correct statement