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.

$0.07 > 0.025$

(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%

Two tailed

(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
	0.30084	(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 1[st] M1: for selection of appropriate model for S 1[st] A1: for a correct values of the parameter p

15 The random variable $Y \sim B(300, 0.6)$.

a Give two reasons why a normal distribution can be used to approximate Y. **(2 marks)**

b Find, using the normal approximation, $P(150 < Y \leqslant 180)$. **(4 marks)**

c Find the largest value of y such that $\underline{P(Y < y) < 0.05}$. **(3 marks)**

a) n is large

p is $\simeq 0.5$

b) $X \sim N(180, 72)$

$P(150 < Y \leq 180)$

$P(150.5 < X < 180.5) = 0.5232$

c) $\underline{\underline{P(Y < y)}} < 0.05$

$P(X < x) < 0.05$

$x = 166.04$

$P(X < 166.04)$

$\underline{\underline{P(Y \leq 166)}}$

$P(Y < 167)$

Hypothesis Testing on the Sample Mean

0	1	2	3	4	5	6	7	8	9

Imagine we have 10 children, one of each age between 0 and 9.
This is our population. There is a **known population mean** of $\mu = 4.5$

					$\bar{x}$
Sample 1:	1	3	7	8	4.75
Sample 2:	6	2	0	9	4.25

...

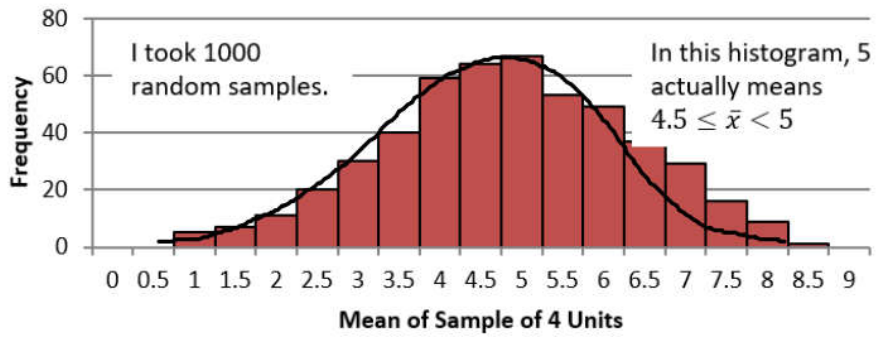
The mean of this sample is $\bar{x} = 4.75$. This sample mean $\bar{x}$ is close the true population mean μ, but is naturally going to vary as we consider different samples.

For a different sample of 4, we might obtain a different sample mean. What would happen if we took lots of different samples of 4, and found the mean $\bar{x}$ of each? How would these means be distributed?

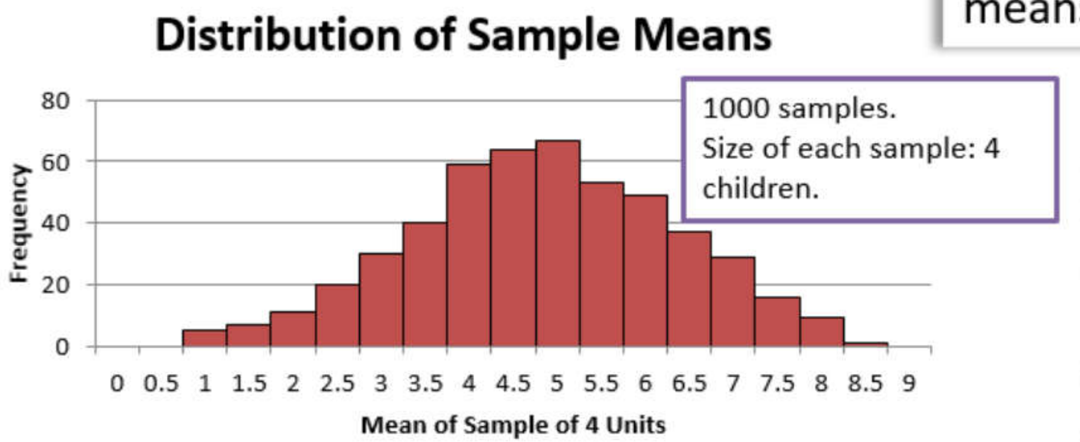
Suppose we took a sample of 4 children.

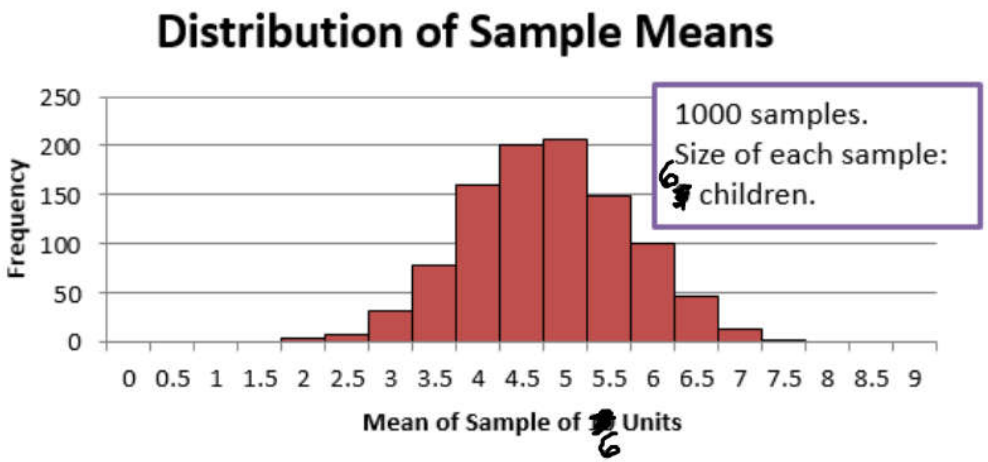
Sample mean $\bar{x}$	**Tally**
4.00	
4.25	|
4.50	
4.75	|
5.00	

Distribution of Sample Means $\bar{X}$



$\bar{X}$ is our distribution across different sample means as we consider different samples.





✎ For a random sample of size n taken from a random variable X, the sample mean $\bar{X}$ is normally distributed with $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Question 1: What type of distribution is $\bar{X}$?
From the left it seems like it is approximately normally distributed!

Question 2: On average, what sample mean do we see? (i.e. the mean of the means!)
μ. The sample means $\bar{x}$ vary around the population mean μ, but on average is μ.

Question 3: Is the variance of $\bar{X}$ (i.e. how spread out the sample means are) the same as that of the variance of the population of children?
No! On the left, we can see that how spread out the sample means are depends on the sample size. If the sample size is small, the sample means are likely to vary quite a bit. But with a larger sample size, we expect the different $\bar{x}$ to be closer to the population mean μ.

A random sample of size n is taken from a population having a normal distribution with mean μ and variance σ^2. Test the hypotheses at the given significance level.

H_0: $\mu = 15$
H_1: $\mu < 15$
$n = 30$
$\sigma = 1.5$
$\bar{y} = 14.5$
1% level

sample mean of my 30 things

$H_0: \mu = 15$

$H_1: \mu < 15$

distribution of all the sample means

$\sigma^2 = \frac{1.5^2}{30}$

$\sigma = \frac{1.5}{\sqrt{30}}$

Assume that H_0 is true, $\bar{Y} \sim N\left(15, \frac{1.5^2}{30}\right)$

$P(\bar{Y} < 14.5) = 0.0339 > 0.01$

Not enough evidence to reject H_0

$\sigma = \frac{12}{5}$

H_0: $\mu = 120$
H_1: $\mu \neq 120$
$n = 25$
$\sigma = 12$
$\bar{x} = 124$
5% level

$\mu > 120$

2.5% two tailed.

$H_0: \mu = 120$

$H_1: \mu \neq 120$

Assume H_0 is true, $\bar{X} \sim N\left(120, \frac{12^2}{25}\right)$

$P(\bar{X} > 124) = 0.0478 > 0.025$

So not enough evidence to reject H_0.

Y

A certain company sells fruit juice in cartons. The amount of juice in a carton has a normal distribution with a standard deviation of 3ml.
The company claims that the mean amount of juice per carton, μ, is 60ml. A trading inspector has received complaints that the company is overstating the mean amount of juice per carton and wishes to investigate this complaint. The trading inspector takes a random sample of 16 cartons and finds that the mean amount of juice per carton is 59.1ml.

Using a 5% level of significance, and stating your hypotheses clearly, test whether or not there is evidence to uphold this complaint.

Let $\bar{Y}$ be the means of the samples.

$H_0: \mu = 60$

$H_1: \mu < 60$

$n = 16$

$\sigma = 3$

$\bar{y} = 59.1$

$\alpha = 5\%$
one tailed

Assume Ho is true

$$\bar{Y} \sim N\left(60, \frac{3^2}{16}\right) \rightarrow \sigma = \frac{3}{4}$$

$$P(\bar{Y} < 59.1) = 0.115 > 0.05$$

Not enough evidence to reject Ho.
There is no evidence to support the claim that the amount of juice is less than 60ml.

Note: Don't confuse X and $\bar{X}$. The X is the distribution over amounts of drink in each individual carton.
$\bar{X}$ is the distribution over **sample means**, i.e. the possible sample means we see as we take samples of 16 cartons. X might not be normally distributed, but $\bar{X}$ will be.

Ex 3G
Q1, 3, 5, 8

Finding the critical region

A random sample of size n is taken from a population having a normal distribution with mean μ and variance σ^2. Find the critical regions for the test statistic $\bar{X}$ for the following levels of significance.

H_0: $\mu = 45$
H_1: $\mu > 45$
$n = 10$
$\sigma = 3$
10% level

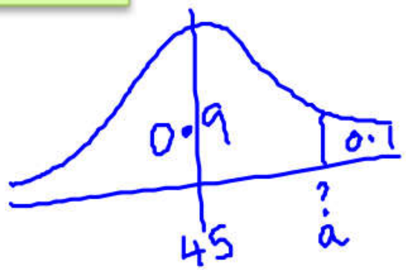
$H_0: \mu = 45$
$H_1: \mu > 45$

Assume H_0 is true, $\bar{X} \sim N\left(45, \frac{3^2}{10}\right)$ → $\sigma = \frac{3}{\sqrt{10}}$

$P(\bar{X} > a) < 0.1$
$P(\bar{X} < a) > 0.9$

$a = 46.2$

Crit region $\bar{X} > 46.2$



H_0: $\mu = 100$
H_1: $\mu \neq 100$
$n = 40$
$\sigma = 15$
5% level

2.5% level

$H_0: \mu = 100$
$H_1: \mu \neq 100$

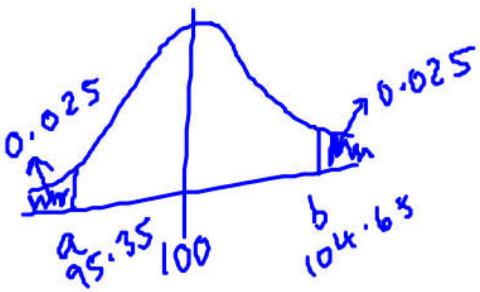
Assume H_0 is true, $\bar{X} \sim N\left(100, \frac{15^2}{40}\right)$ → $\sigma = \frac{15}{\sqrt{40}}$

$P(\bar{X} < a) = 0.025$
$a = 95.35$

Crit region $\bar{X} < 95.35$

$P(\bar{X} < b) = 0.975$
$b = 104.65$

Crit region $\bar{X} > 104.65$



A machine products bolts of diameter D where D has a normal distribution with mean 0.580 cm and standard deviation 0.015 cm. The machine is serviced and after the service a random sample of 50 bolts from the next production run is taken to see if the mean diameter of the bolts has changed from 0.580 cm. The distribution of the diameters of bolts after the service is still normal with a standard deviation of 0.015 cm.

(a) Find, at the 1% level, the critical region for this test, stating your hypotheses clearly.
The mean diameter of the sample of 50 bolts is calculated to be 0.587 cm.
(b) Comment on this observation in light of the critical region.

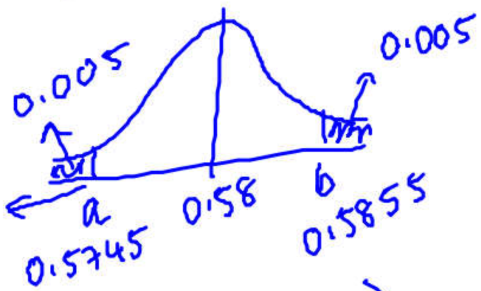
$\mu = 0.58$
$\sigma = 0.015$
$n = 50$
$\alpha = 1\%$
0.5%
0.005

a) $H_0: \mu = 0.58$
$H_1: \mu \neq 0.58$

$\sigma = \dfrac{0.015}{\sqrt{50}}$

Assume H_0 is true, $\bar{D} \sim N\left(0.58, \dfrac{0.015^2}{50}\right)$

$P(\bar{D} < a) = 0.005$
$a = 0.5745$ Crit region $\bar{D} < 0.5745$



$P(\bar{D} < b) = 0.995$
$b = 0.5855$ Crit region $\bar{D} > 0.5855$.

b) $0.587 > 0.5855$ and so is in the critical region.
✓ There is evidence to reject H_0.
✓ This supports that the mean diameter has changed.

Ex 3G
Q2, 4

Ex 3G
Q6, 7

3. A machine cuts strips of metal to length L cm, where L is normally distributed with standard deviation 0.5 cm. $\sigma = 0.5$

Strips with length either less than 49 cm or greater than 50.75 cm **cannot** be used.

Given that 2.5% of the cut lengths exceed 50.98 cm, $\mu = ?$

(a) find the probability that a randomly chosen strip of metal **can** be used. 0.9104 **(5)**

Ten strips of metal are selected at random.

(b) Find the probability fewer than 4 of these strips **cannot** be used. 0.0896 **(2)**

A second machine cuts strips of metal of length X cm, where X is normally distributed with standard deviation 0.6 cm

A random sample of 15 strips cut by this second machine was found to have a mean length of 50.4 cm

(c) Stating your hypotheses clearly and using a 1% level of significance, test whether or not the mean length of all the strips, cut by the second machine, is greater than 50.1 cm **(5)**

a)
$P(L > 50.98) = 0.025$
$P(L < 50.98) = 0.975$
$P(Z < a) = 0.975$
1.96
$\dfrac{50.98 - \mu}{0.5} = 1.96$
$\mu = 50$

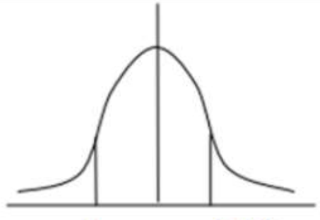
$P(49 < L < 50.75) = \underline{0.9104}$

b) $X \sim B(10, 0.0896)$
$P(X \leq 3) = \underline{0.9113}$

c) $\sigma = 0.6$
$n = 15$
$\bar{y} = 50.4$
$\alpha = 1\%$

$H_0: \mu = 50.1$
$H_1: \mu > 50.1$
$\bar{Y} \sim N\left(50.1, \dfrac{0.6^2}{15}\right) \rightarrow \sigma = \dfrac{0.6}{\sqrt{15}}$
$P(\bar{Y} > 50.4) = 0.0264 > 0.01$

So not enough evidence to reject H_0, no reason to believe mean is greater than 50.1 cm

Question	Scheme	Marks	AOs
Q3(a)			
	P $(L > 50.98) = 0.025$	B1cao	
	$\therefore \ \dfrac{50.98-\mu}{0.5} = 1.96$	M1	
	$\therefore \ \mu = 50$	A1cao	
	$\mathrm{P}(49 < L < 50.75)$	M1	
	$= 0.9104\ldots$ awrt **<u>0.910</u>**	A1ft	
		(5)	
(b)	S = number of strips that cannot be used so $S \sim \mathrm{B}(10, 0.090)$	M1	3.3
	$= \mathrm{P}(S \leqslant 3) = 0.991166\ldots$ awrt 0.991	A1	1.1b
		(2)	
(c)	$\mathrm{H}_0 : \mu = 50.1$ $\quad \mathrm{H}_1 : \mu > 50.1$	B1	2.5
	$\bar{X} \sim \mathrm{N}\left(50.1, \dfrac{0.6^2}{15}\right)$ and $\bar{X} > 50.4$	M1	3.3
	$\mathrm{P}(\bar{X} > 50.4) = 0.0264$	A1	3.4
	$p = 0.0264 > 0.01$ <u>or</u> $z = 1.936\ldots < 2.3263$ and not significant	A1	1.1b
	There is insufficient evidence that the **<u>mean length</u>** of strips is **<u>greater than 50.1</u>**	A1	2.2b
		(5)	
		(12 marks)	

Question 3 continued

Notes:

(a)

1st M1: for standardizing with μ and 0.5 and setting equal to a z value ($|z| > 1$)

2nd M1: for attempting the correct probability for strips that can be used

2nd A1ft: awrt 0.910 (allow ft of their μ)

(b)

M1: for identifying a suitable binomial distribution

A1: awrt 0.991 (from calculator)

(c)

B1: hypotheses stated correctly

M1: for selecting a correct model (stated or implied)

1st A1: for use of the correct model to find p = awrt 0.0264 (allow z = awrt 1.94)

2nd A1: for a correct calculation, comparison and correct statement

3rd A1: for a correct conclusion in context mentioning "mean length" and 50.1

5. A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, D ml, follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than 24.63 ml

(a) find, to 2 decimal places, the value of k such that $P(24.63 < D < k) = 0.45$

(5)

A random sample of 200 bottles is taken.

(b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml

(3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml

(c) Test Hannah's belief at the 5% level of significance. You should state your hypotheses clearly.

(5)

Part	Working or answer an examiner might expect to see	Mark	Notes
(a)	$\dfrac{24.63-25}{\sigma}=-1.0364$	M1	This mark is given for standardising as part of a method to find σ
	$\sigma=0.357$	A1	This mark is given for a correct value of σ
	$\mathrm{P}(D>K)=0.4$ or $\mathrm{P}(D<K)=0.6$	B1	This mark is given for
	$\dfrac{k-25}{\sigma}=\dfrac{k-25}{0.357}=0.2533$	M1	This mark is given for using a normal model to find the probability
	$k=25.09$	A1	This mark is given for a correct value for k
(b)	$Y\sim \mathrm{B}(200,\ 0.45)$ so $W\sim \mathrm{N}(90,\ 49.5)$	B1	This mark is given for setting up the normal distribution approximation of the binomial
	$\mathrm{P}(Y<100)\approx \mathrm{P}(W<99.5)=\mathrm{P}\left(Z<\dfrac{99.5-90}{\sqrt{49.5}}\right)$	M1	This mark is given for using the normal model with a continuity correction
	$=0.912$	A1	This mark is given for finding a correct value of the probability
(c)	$\mathrm{H}_0: \mu=25$ $\mathrm{H}_1: \mu<25$	B1	This mark is given for both hypotheses in terms of μ found correctly
	$\overline{D}\sim \mathrm{N}\left(25, \dfrac{0.16^2}{20}\right)$	M1	This mark is given for a method to set up the normal distribution
	$\mathrm{P}(\overline{D}<24.94)=0.0468$	A1	This mark si govern for using the model to find a correct p-value
	$p=0.0468<0.05$, so reject H_0	M1	This mark si given for a correct comparison and non-contextual conclusion
	There is sufficient evidence to support Hannah's belief	A1	This mark is given for a correct conclusion in context stated