Normal Distributions and Hypothesis Testing

1:: Characteristics of the Normal Distribution

What shape is it? What parameters does it have?

2:: Finding probabilities on a standard normal curve.

"Given that IQ is distributed as $X \sim N(100, 15^2)$, determine the probability that a randomly chosen person has an IQ above 130."

3:: Finding unknown means/standard deviations.

In Wales, 30% of people have a height above 1.6m. Given the mean height is 1.4m and heights are normally distributed, determine the standard deviation of heights.

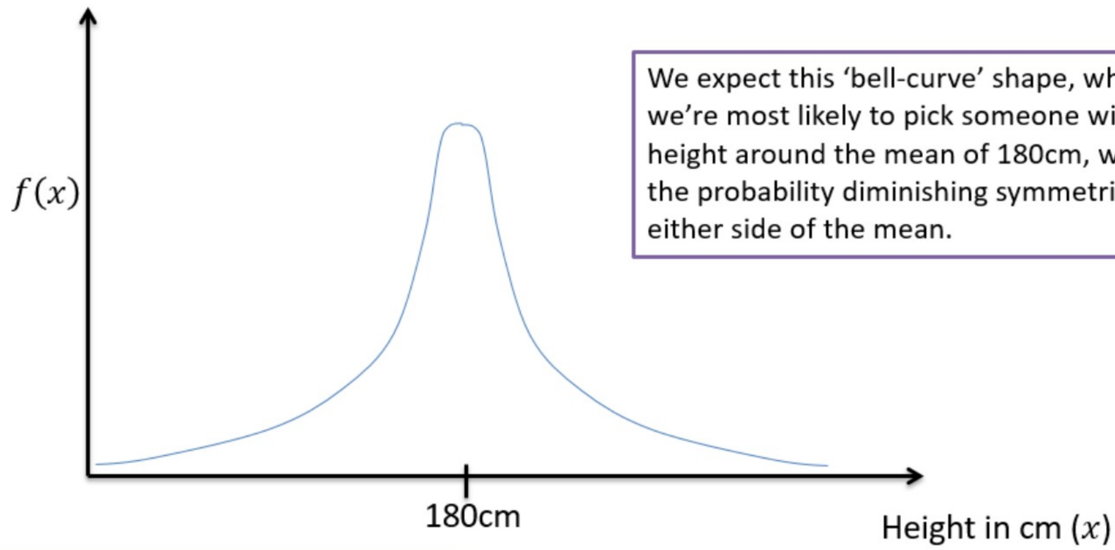
4:: Binomial → Normal Approximations

How would I approximate $X \sim B(10, 0.4)$ using a Normal distribution? Under what conditions can we make such an approximation?

5:: Hypothesis Testing

What does it look like?

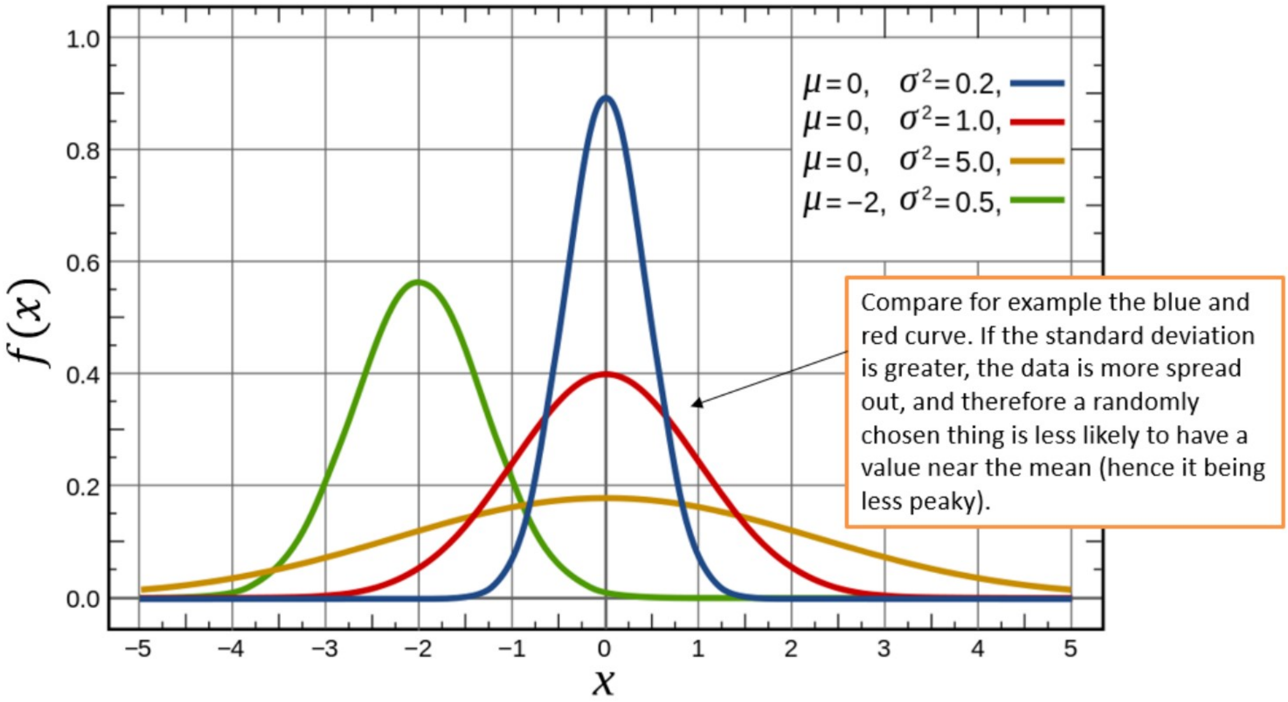
The following shows what the probability distribution might look like for a random variable X, if X is the height of a randomly chosen person.



We expect this 'bell-curve' shape, where we're most likely to pick someone with a height around the mean of 180cm, with the probability diminishing symmetrically either side of the mean.

A variable with this kind of distribution is said to have a **normal distribution**.

For normal distributions we tend to draw the y axis at the mean for symmetry.



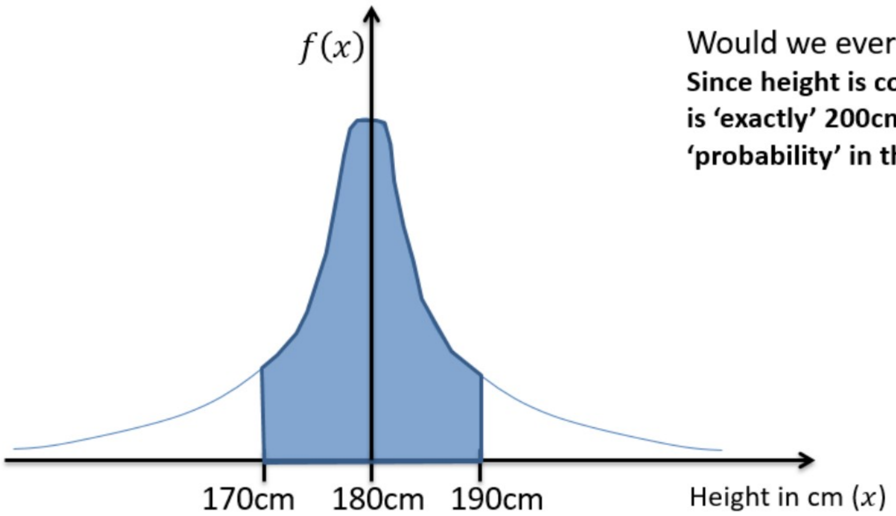
We can set the mean μ and the standard deviation σ of the Normal Distribution. If a random variable X is normally distributed, then we write

$$\boldsymbol{X \sim N(\mu, \sigma^2)}$$

Normal Distribution Facts

For a Normal Distribution to be used, the variable has to be:
continuous

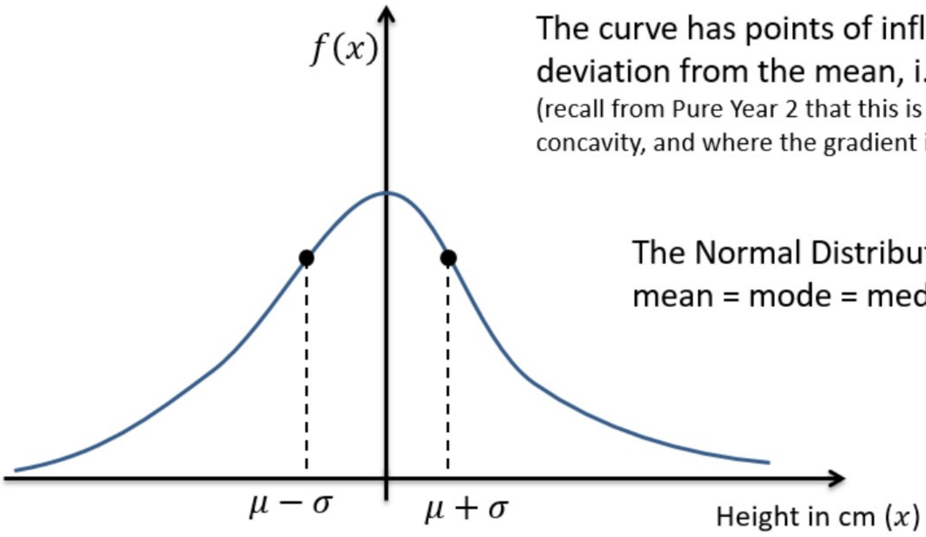
To find $P(170 < X < 190)$, we could:
find the area between these values.



With a discrete variable, all the probabilities had to add up to 1.
For a continuous variable, similarly:
the area under the probability graph has to be 1.

Would we ever want to find $P(X = 200)$ say?
Since height is continuous, the probability someone is 'exactly' 200cm is infinitesimally small. So not a 'probability' in the normal sense.

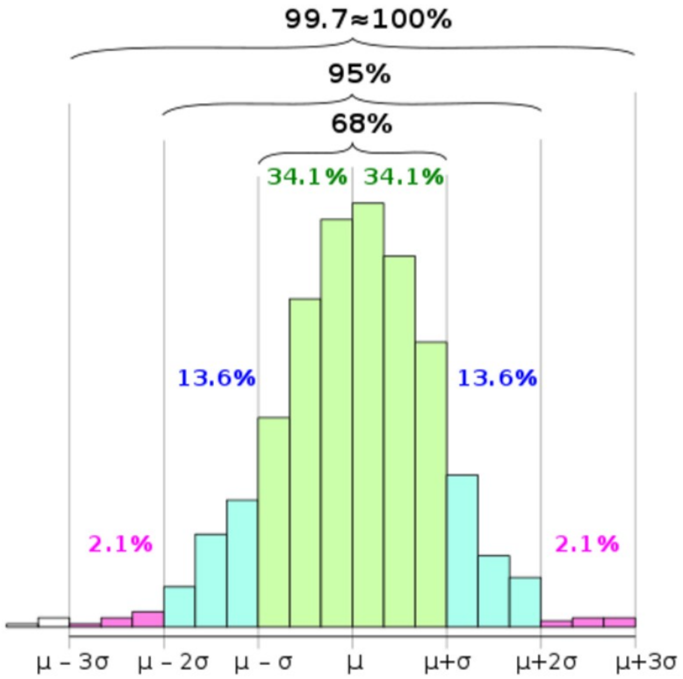
Side Notes: You might therefore wonder what the y-axis actually is. It is **probability density**, i.e. "the probability per unit cm". This is analogous to frequency density with histograms, where the y-value is frequency density area under the graph gives frequency. We use $f(x)$ rather than $p(x)$, to indicate probability density.



The curve has points of inflection one standard deviation from the mean, i.e. $\mu \pm \sigma$

(recall from Pure Year 2 that this is where the curve changes concavity, and where the gradient is not changing)

The Normal Distribution is symmetrical, i.e. mean = mode = median



The histogram above is for a quantity which is approximately normally distributed.

The 68-95-99.7 rule is a shorthand used to remember the percentage of data that is within 1, 2 and 3 standard deviations from the mean respectively.

<u>You need to memorise this!</u>

✎

$\approx 68\%$ of data is within one standard deviation of the mean.

$\approx 95\%$ of data is within two standard deviations of the mean.

$\approx 99.7\%$ of data is within three standard deviations of the mean.

For practical purposes we consider all data to lie within $\mu \pm 5\sigma$

Only one in 1.7 million values fall outside $\mu \pm 5\sigma$. CERN used a "5 sigma level of significance" to ensure the data suggesting existence of the Higgs Boson wasn't by chance: this is a 1 in 3.5 million chance (if we consider just one tail).

The diameters of a rivet produced by a particular machine, X mm, is modelled as $X \sim N(8, 0.2^2)$. Find:

a) $P(X > 8)$

b) $P(7.8 < X < 8.2)$

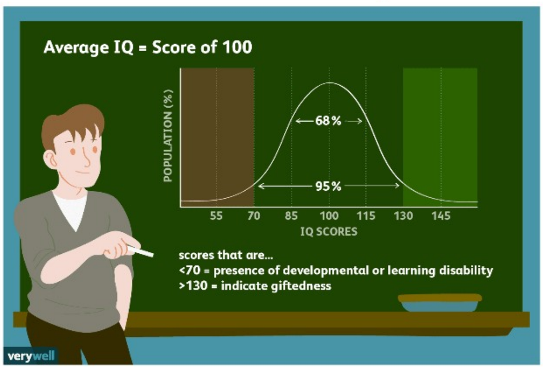
Tip: Draw a diagram!

IQ ("Intelligence Quotient") for a given population is, by definition, distributed using $X \sim N(100, 15^2)$. Find:

a) $P(70 < X < 130)$

b) $P(X > 115)$

Tip: Draw a diagram!



Your Turn

4 The armspans of a group of Year 5 pupils, X cm, are modelled as $X \sim N(120, 16)$.

a State the proportion of pupils that have an armspan between 116 cm and 124 cm.

b State the proportion of pupils that have an armspan between 112 cm and 128 cm.

5 The lengths of a colony of adders, Y cm, are modelled as $Y \sim N(100, \sigma^2)$. If 68% of the adders have a length between 93 cm and 107 cm, find σ^2.

7 The masses of the pigs, M kg, on a farm are modelled as $M \sim N(\mu, \sigma^2)$. If 84% of the pigs weigh more than 52 kg and 97.5% of the pigs weigh more than 47.5 kg, find μ and σ^2.

Getting normal values from your calculator

IQ is distributed using $X \sim N(100, 15^2)$. Find

(a) $P(X < 109)$

(b) $P(X \geq 93)$

(c) $P(110 < X < 120)$

(d) $P(X < 80 \; or \; X > 106)$

Please: draw a diagram!

3 The random variable $X \sim N(25, 25)$.

Find: **a** $P(Y < 20)$ **b** $P(18 < Y < 26)$ **c** $P(Y > 23.8)$

4 The random variable $X \sim N(18, 10)$.

Find: **a** $P(X \geqslant 20)$ **b** $P(X < 15)$ **c** $P(18.4 < X < 18.7)$

6 The random variable $T \sim N(4.5, 0.4)$.

a Find $P(T < 4.2)$.

b Without further calculation, write down $P(T > 4.2)$.

7 The random variable $Y \sim N(45, 2^2)$. Find:

a $P(Y < 41 \text{ or } Y > 47)$ **b** $P(Y < 44 \text{ or } 46.5 < Y < 47.5)$

3	a	0.1587	b 0.4985	c 0.5948
4	a	0.2635	b 0.1714	c 0.0373
5	a	i 0.7475	ii 0.2525	
	b	Sum is 1, combined probabilities include every possible value.		
6	a	0.3176	b 0.6824	
7	a	0.1814	b 0.4295	

Using normal probabilities in questions

The criteria for joining Mensa is an IQ of at least 131.
Assuming that IQ has the distribution $X \sim N(100, 15^2)$ for a population, determine:

a) What percentage of people are eligible to join Mensa.

b) If 30 adults are randomly chosen, the probability that at least 3 of them will be eligible to join.

Inverse Normal Distribution

We now know how to use a calculator to value of the variable to obtain a probability. But we might want to do the reverse: given a probability of being in a region, how do we find the value of the boundary?

$X \sim N(20,3^2)$. Find, correct to two decimal places, the values of a such that:

a. $P(X < a) = 0.75$
b. $P(X > a) = 0.4$
c. $P(16 < X < a) = 0.3$

DRAW A SKETCH!

The 'area' requested by your calculator is the probability **up to** the value of interest (in this case a)

The graphics calculator is more advanced – you can assign whether the probability tails to the left (up to the value *a*), to the right (above the value *a*) or is symmetrically in the centre.

If the IQ of a population is distributed using $X \sim N(100,15^2)$.

a. Determine the IQ corresponding to the top 30% of the population.
b. Determine the interquartile range of IQs.

In general the quartiles of a normal distribution are approximately $\mu \pm \frac{2}{3}\sigma$

$X \sim N(80, 7^2)$. Using your calculator,

a. determine the a such that $P(X > a) = 0.65$
b. determine the b such that $P(75 < X < b) = 0.4$
c. determine the c such that $P(c < X < 76) = 0.2$
d. determine the interquartile range of X.

Ex 3C

Standard Normal Distribution

✏ Z is the number of standard deviations above the mean.

If again we use IQ distributed as $X \sim N(100, 15^2)$ then: (in your head!)

IQ	Z
100	0
130	2
85	-1
165	4.333
62.5	-2.5

This formula makes sense if you think about the definition above. For an IQ of 130: $Z = \frac{130-100}{15} = 2$ as expected.

✏ Z represents the coding:

$$Z = \frac{X - \mu}{\sigma}$$

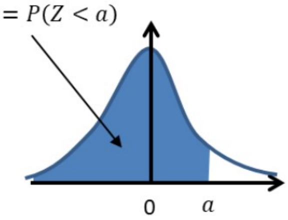
and $Z \sim N(0, 1^2)$. Z is known as a **standard** normal distribution.

The 0 and 1 of $Z \sim N(0, 1^2)$ are the result of the coding. If we've subtracted μ from each value the mean of the normal distribution is now 0. If we've divided all the values by σ the standard deviation is now $\frac{\sigma}{\sigma} = 1$

The point of coding in this context is that all different possible normal distributions become a single unified distribution where we no longer have to worry about the mean and standard deviation.
It means for example when we calculate $P(Z < 3)$, this will always give the same probability regardless of the original distribution.

It also means we can look up probabilities in a $\boldsymbol{z}$**-table**:

$\phi(a) = P(Z < a)$



✎ $\Phi(a) = P(Z < a)$ is the cumulative distribution for the standard normal distribution. The values of $\Phi(a)$ can be found in a z-table.

This is a traditional z-table in the old A Level syllabus (but also found elsewhere). You no longer get given this and are expected to use your calculator.

This is from the new formula booklet. This is sometimes known as a 'reverse z-table', because you're looking up the z-value for a probability. Beware: p here it the probability of **exceeding** z rather than being up to z. Let's use it...

THE NORMAL DISTRIBUTION FUNCTION

The function tabulated below is $\Phi(z)$, defined as $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\,dt$.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.02	0.9783
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.04	0.9793
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.06	0.9803
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.08	0.9812
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.10	0.9821
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.12	0.9830
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.14	0.9838
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.16	0.9846
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.18	0.9854
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.20	0.9861

Percentage Points of The Normal Distribution

The values z in the table are those which a random variable $Z - N(0, 1)$ exceeds with probability p; that is, $P(Z > z) = 1 - \Phi(z) = p$.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

The random variable $X \sim N(50, 4^2)$. Write in terms of $\Phi(z)$ for some value of z.
(a) $P(X < 53)$ (b) $P(X \geq 55)$

The systolic blood pressure of an adult population, S mmHg, is modelled as a normal distribution with mean 127 and standard deviation 16. A medical research wants to study adults with blood pressures higher than the 95th percentile. Find the minimum blood pressure for an adult included in her study.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

(a) Determine $P(Z > -1.3)$
(b) Determine $P(-2 < Z < 1)$
(c) Determine the a such that $P(Z > a) = 0.7$
(d) Determine the a such that $P(-a < Z < a) = 0.6$

The aim here is to rewrite using standardised form first – this will be helpful in the next few exercises, trust me!

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

Tip: Either changing < to/from > or changing the sign (+ to/from -) has the effect of "$1 -$". However, if you change both, the "$1 -$"s cancel out!

$$P(a < Z < b) = P(Z < b) - P(Z < a)$$

IQ is distributed with mean 100 and standard deviation 15. Using an appropriate table, determine the IQ corresponding to the
(a) top 10% of people.
(b) bottom 20% of people.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

If $X \sim N(100, 15^2)$, determine, in terms of Φ:
(a) $P(X > 115)$
(b) $P(77.5 < X < 112)$

Find the a such that:
(a) $P(-a < Z < a) = 0.2$
(b) $P(0 < Z < a) = 0.35$

~~Ex 3D~~

Missing μ and σ

In the last section, you may have thought, "what's the point of standardising to Z when I can just use the DISTRIBUTION mode on my calculator?"
Fair point, but both forward and reverse normal lookups on the calculator **required you to specify μ and σ.**

$X \sim N(\mu, 3^2)$. Given that $P(X > 20) = 0.2$, find the value of μ.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

A machine makes metal sheets with width, X cm, modelled as a normal distribution such that $X \sim N(50, \sigma^2)$.
(a) Given that $P(X < 46) = 0.2119$, find the value of σ.
(b) Find the 90th percentile of the widths.

The method here is exactly the same as before:
1. Using a sketch, determine whether you're expecting a positive or negative z value.
2. Look up z value, using tables if you can (otherwise your calculator). Make negative if in bottom half.
3. Use $Z = \frac{X-\mu}{\sigma}$

When both are missing

If both μ and σ are missing, we end up with simultaneous equations which we must solve.

Edexcel S1 Jan 2011

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.

(c) Given that $\mathrm{P}(Y < 160) = 0.99$ and $\mathrm{P}(Y > 152) = 0.90$, find the value of μ and the value of σ. **(6)**

4. A company has a customer services call centre. The company believes that the time taken to complete a call to the call centre may be modelled by a normal distribution with mean 16 minutes and standard deviation σ minutes.

Given that 10% of the calls take longer than 22 minutes,

(a) show that, to 3 significant figures, the value of σ is 4.68 (3)

(b) Calculate the percentage of calls that take less than 13 minutes. (1)

A supervisor in the call centre claims that the mean call time is less than 16 minutes. He collects data on his own call times.

- 20% of the supervisor's calls take more than 17 minutes to complete.
- 10% of the supervisor's calls take less than 8 minutes to complete.

Assuming that the time the supervisor takes to complete a call may be modelled by a normal distribution,

(c) estimate the mean and the standard deviation of the time taken by the supervisor to complete a call. (6)

(d) State, giving a reason, whether or not the calculations in part (c) support the supervisor's claim. (1)

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)**

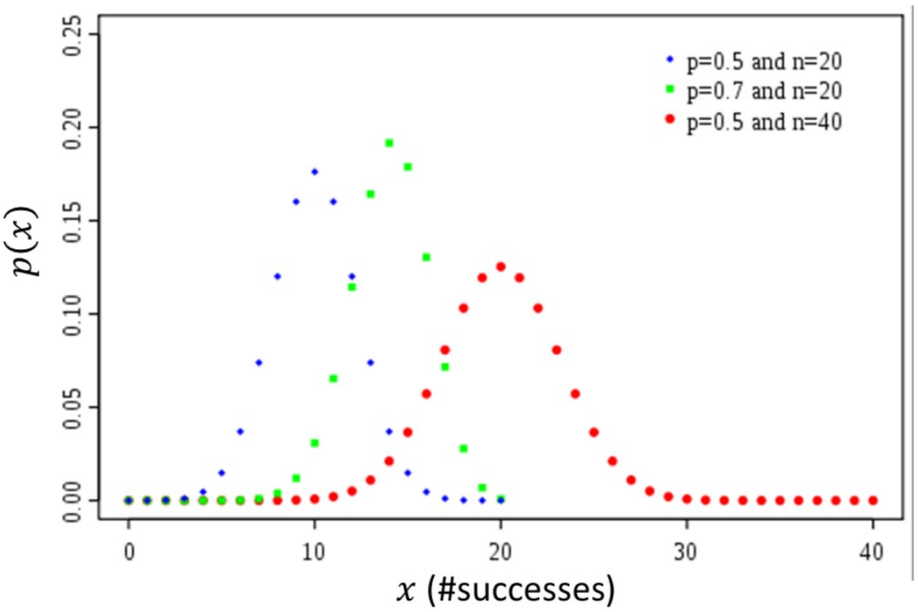
(*b*) Obtain a second equation in μ and σ. **(3)**

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

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

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

✏ 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:

$X \sim B(10, 0.2) \quad \rightarrow \quad Y \sim N($

$X \sim B(20, 0.5) \quad \rightarrow \quad Y \sim N($

$X \sim B(6, 0.3) \quad \rightarrow \quad Y \sim N($

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 $P(X \leq 65)$. ☹

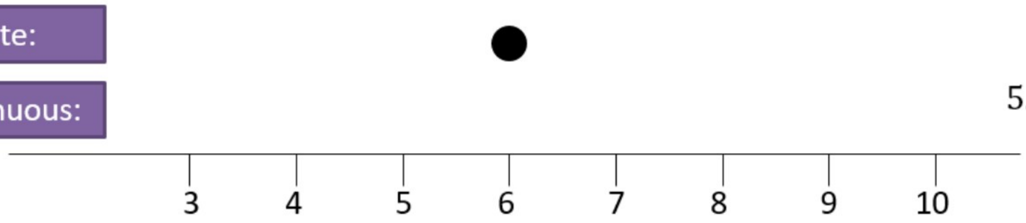
Continuity Corrections

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)?

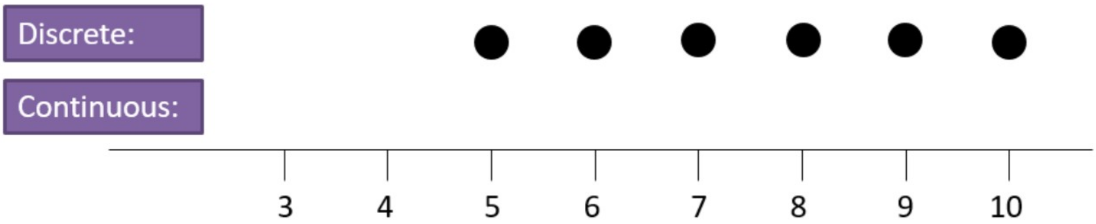
Discrete: $X = 6$

Continuous: $5.5 < Y < 6.5$



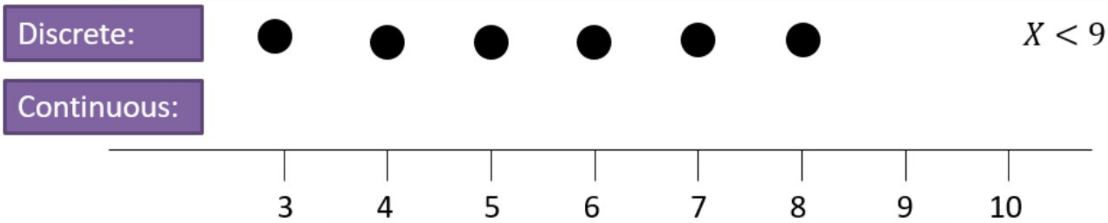
We can't just find $P(Y = 6)$ when Y is continuous, because the probability is effectively 0. But $P(5.5 < Y < 6.5)$ would seem a sensible interval to use because any time between 5.5 and 6.5 would have rounded to 6 hours were it discrete.

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



$X \geq 5$

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



$X < 9 \rightarrow X \leq 8$

✎ A continuity correction is approximating a discrete range using a continuous one.
1. If > or <, convert to ≥, ≤ first.
2. Enlarge the range by 0.5.

Discrete	➡ Continuous
$P(X \leq 7)$	
$P(X < 10)$	
$P(X > 9)$	
$P(1 \leq X \leq 10)$	
$P(3 < X < 6)$	
$P(3 \leq X < 6)$	
$P(3 < X \leq 6)$	
$P(X = 3)$	

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

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

Edexcel S2 Jan 2004 Q3

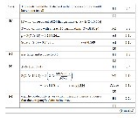
The discrete random variable X is distributed $B(n, p)$.

(*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 $P(90 \leq X < 105)$. **(7)**

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.

(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.

(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.

(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)

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,

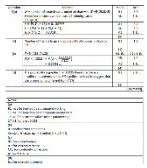
(c) use a normal approximation to find the probability that at most 122 of these 300 scratchcards are winning scratchcards.

(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)



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

…

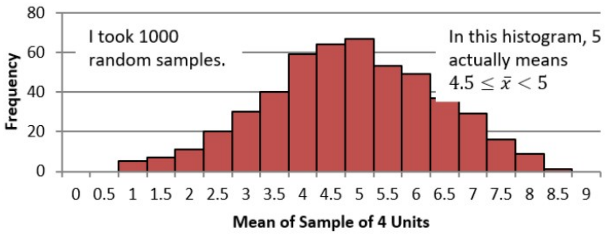
The mean of this sample is $\bar{x} = 4.75$. This sample mean $\bar{x}$ is close the true population mean μ, but is naturally going to vary as we consider different samples.

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.

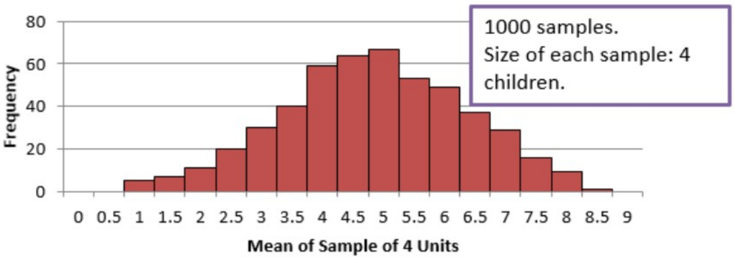
Sample mean $\bar{x}$	Tally
4.00	
4.25	|
4.50	
4.75	|
5.00	

Distribution of Sample Means $\bar{X}$

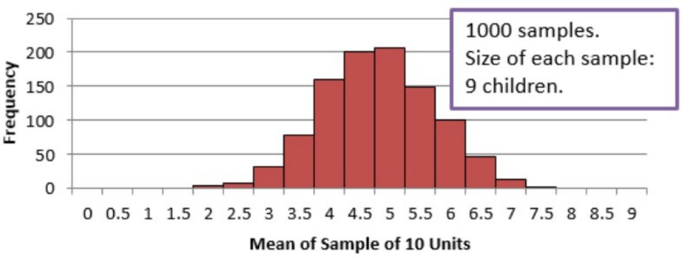


$\bar{X}$ is our distribution across different sample means as we consider different samples.

Distribution of Sample Means



Distribution of Sample Means



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

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

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.

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

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.

Ex 3G
Q2, 4

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

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,

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

Ten strips of metal are selected at random.

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

Ex 3G
Q6, 7

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)

Conditional Probabilities

This is not in the textbook. But given the recent Chapter 2 on Conditional Probabilities and the fact that the type of question below occurred frequently in old S1 papers, it seems worthwhile to cover!

Edexcel S1 May 2014(R) Q4

The time, X minutes, taken to fly from London to another city has a normal distribution with mean μ minutes.

Given that $\mathrm{P}(X < \mu - 15) = 0.35$

(*c*) find $\mathrm{P}(X > \mu + 15 \mid X > \mu - 15)$. **(3)**

Edexcel S1 Jan 2013 Q4a,c

The length of time, L hours, that a phone will work before it needs charging is normally distributed with a mean of 100 hours and a standard deviation of 15 hours.

(*a*) Find $\mathrm{P}(L > 127)$.

(3)

Alice is about to go on a 6 hour journey. Given that it is 127 hours since Alice last charged her phone,

(*c*) find the probability that her phone will not need charging before her journey is completed.

(4)