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.

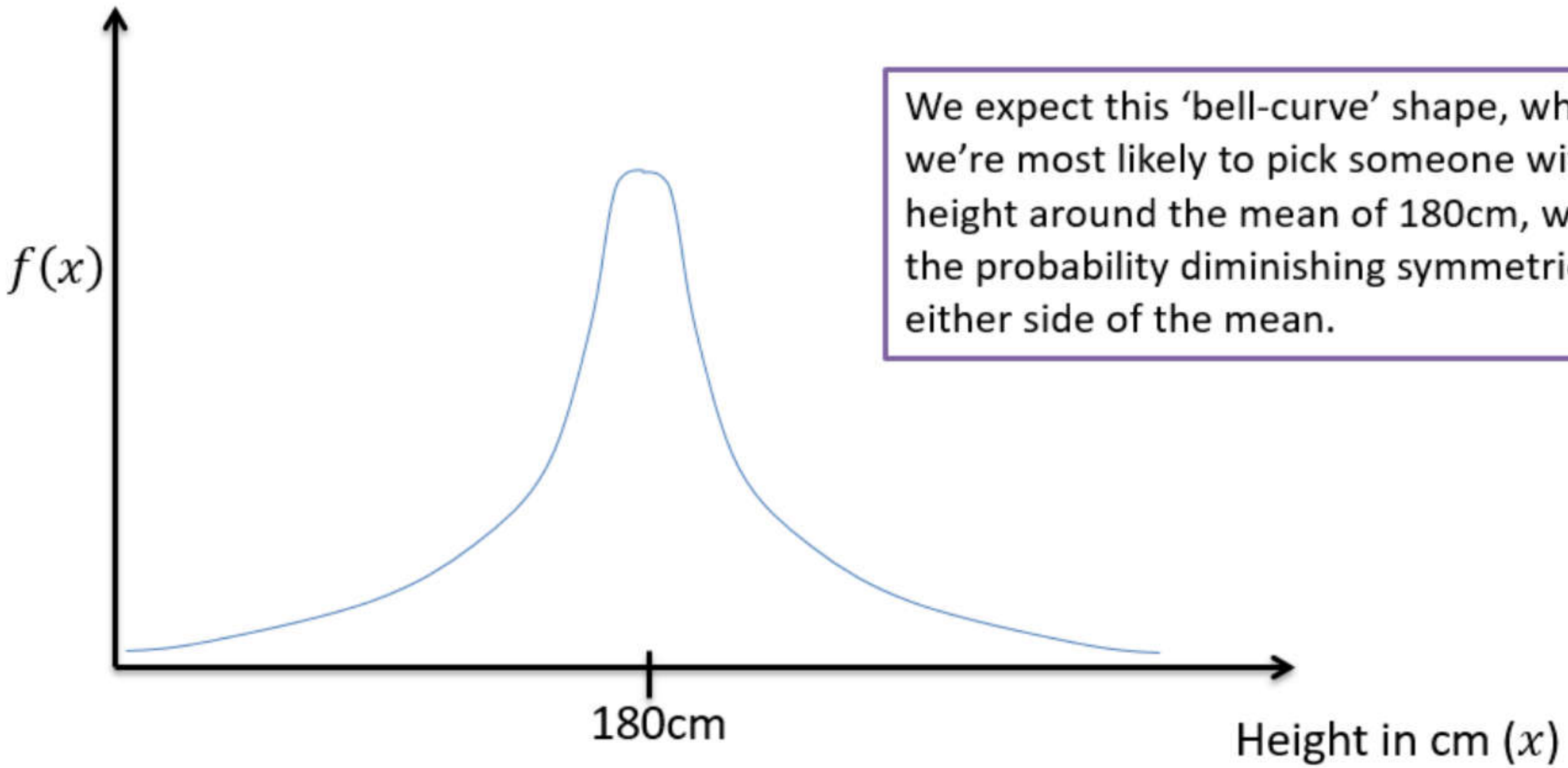
4:: Binomial → Normal Approximations

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?

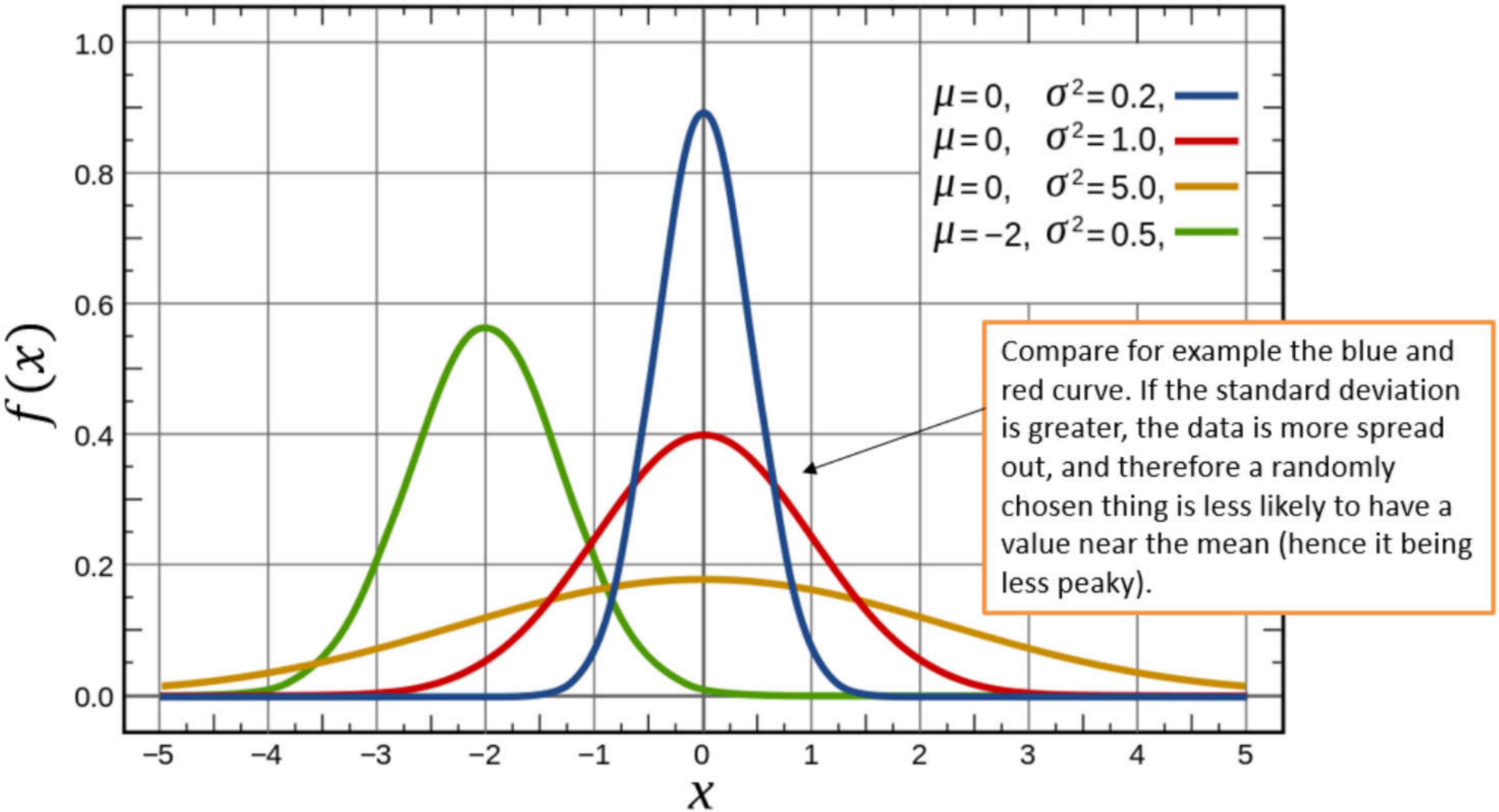
The following shows what the probability distribution might look like for a random variable X, if X is the height of a randomly chosen person.



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

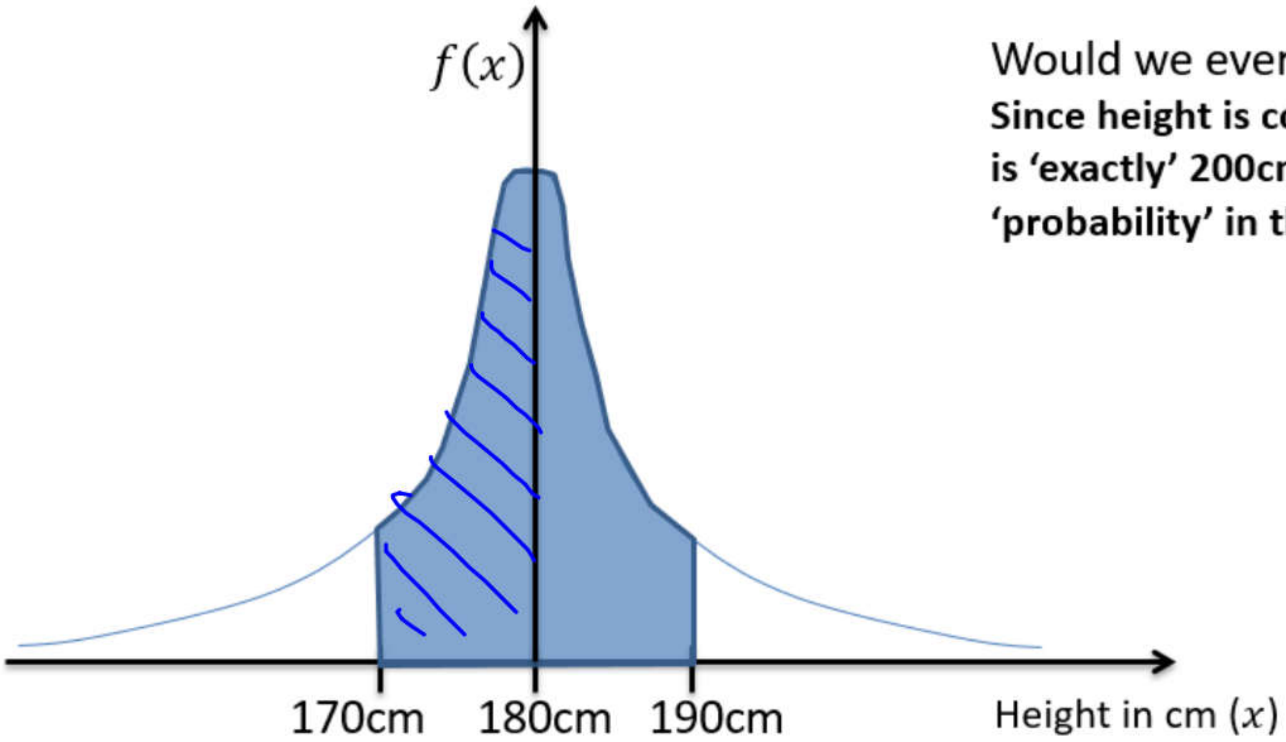
$$\boldsymbol{X \sim N(\mu, \sigma^2)}$$

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

Normal Distribution Facts

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

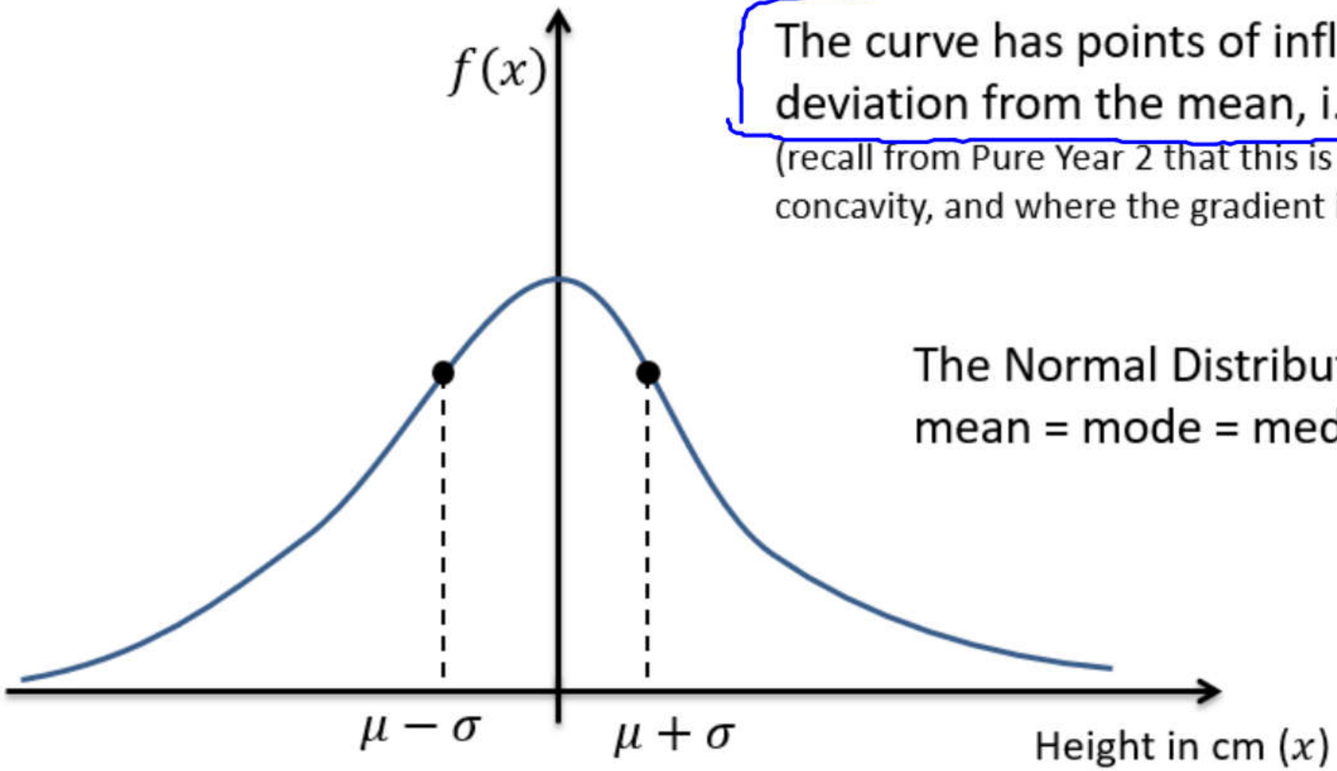
To find $P(170 < X < 190)$, we could:
find the area between these values.



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.

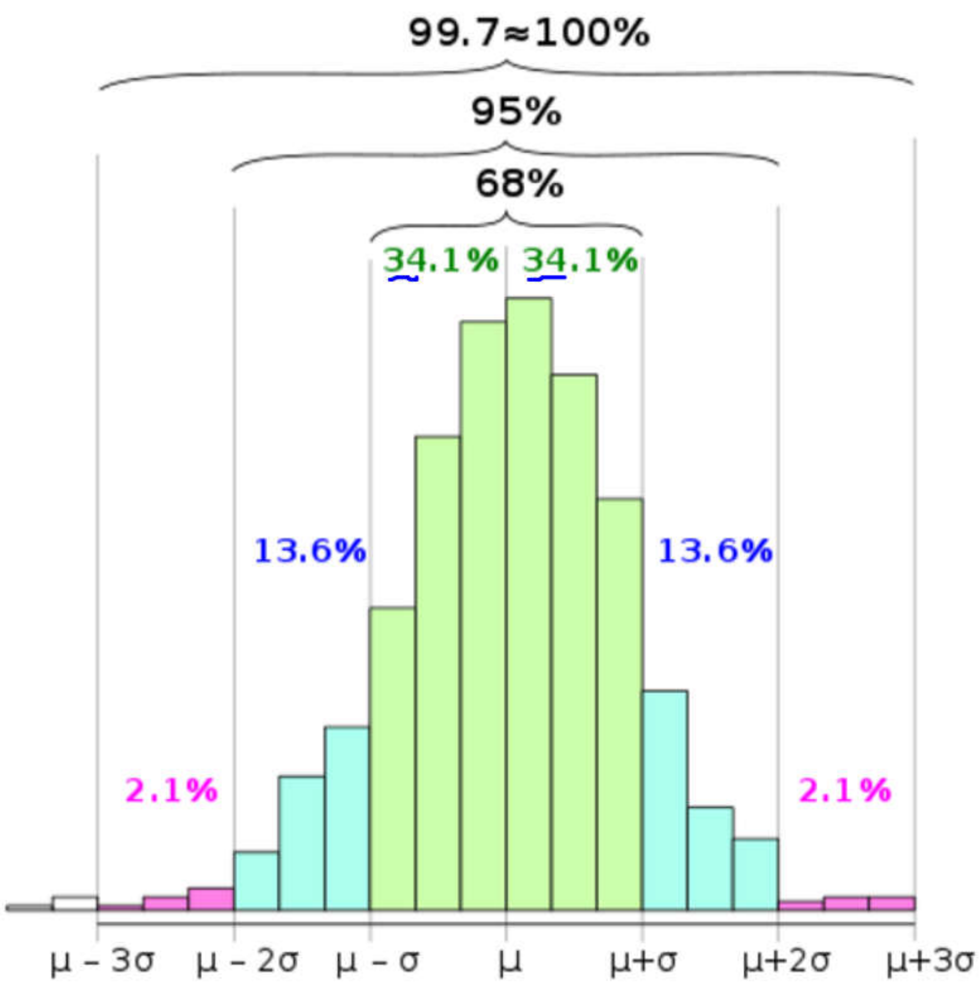
Side Notes: You might therefore wonder what the y-axis actually is. It is **probability density**, i.e. "the probability per unit cm". This is analogous to frequency density with histograms, where the y-value is frequency density area under the graph gives frequency. We use $f(x)$ rather than $p(x)$,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

You need to memorise this!

≈ 68% of data is within one standard deviation of the mean.
≈ 95% of data is within two standard deviations of the mean.
≈ 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).

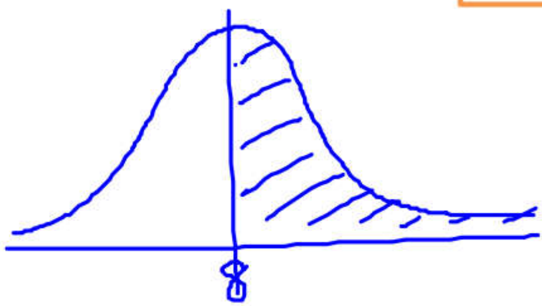
The diameters of a rivet produced by a particular machine, X mm, is modelled as $X \sim N(8, 0.2^2)$. Find:

a) $P(X > 8)$

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

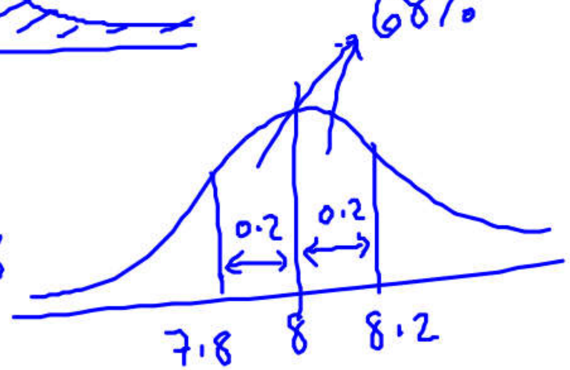
Tip: Draw a diagram!

$X \sim N(8, 0.2^2)$



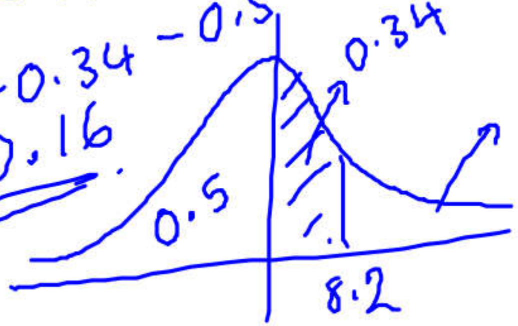
a) $P(X > 8) = 0.5$

b) $P(7.8 < X < 8.2) = 0.68$



$\sigma = 0.2$

c) $P(X > 8.2) = 1 - 0.34 - 0.5$
$= \underline{0.16}$



$X \sim N(8, 0.01)$

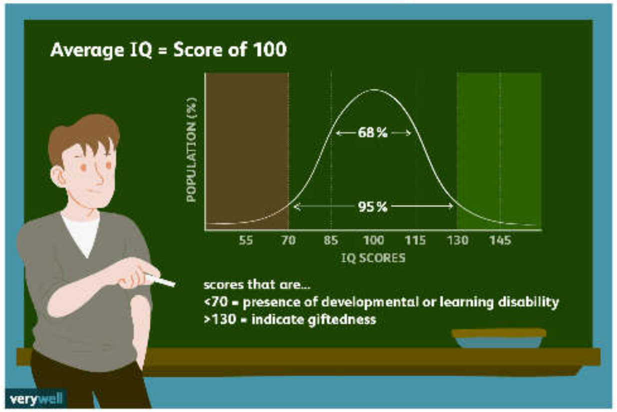
$\sigma = \sqrt{0.01}$
$= 0.1$

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

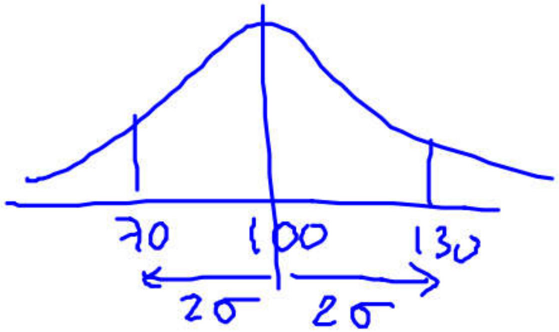
a) $P(70 < X < 130)$

b) $P(X > 115)$

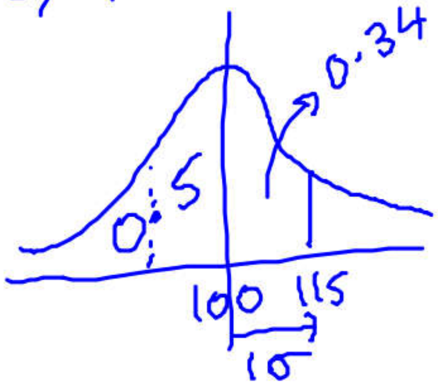
Tip: Draw a diagram!



a) $P(70 < X < 130) = 0.95$



b) $P(X > 115) = 1 - 0.5 - 0.34 = \underline{0.16}$.



Your Turn

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

$\mu = 120$
$\sigma = 4$

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.

a) 0.68

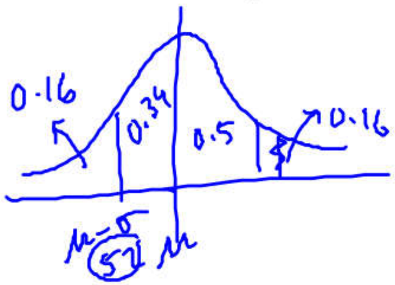
b) 0.95

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.

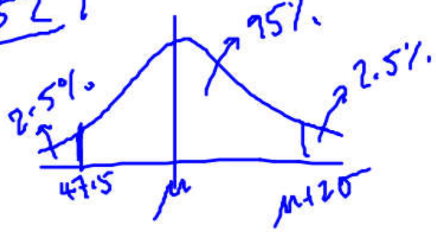
$107 - 100 = 7$
$100 - 93 = 7$

$\sigma = 7$
$\sigma^2 = 49.$

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.



$\boxed{\mu - \sigma = 52}$



$\boxed{\mu - 2\sigma = 47.5}$

$\mu = 56.5$
$\sigma = 4.5$

Getting normal values from your calculator

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

(a) $P(X < 109)$

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

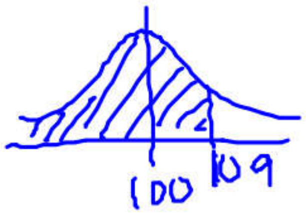
(c) $P(110 < X < 120)$

(d) $P(X < 80 \text{ or } X > 106)$

Please: draw a diagram!

$\mu = 100 \quad \sigma = 15$

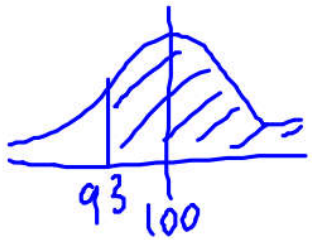
a) $P(X < 109) = 0.7257$



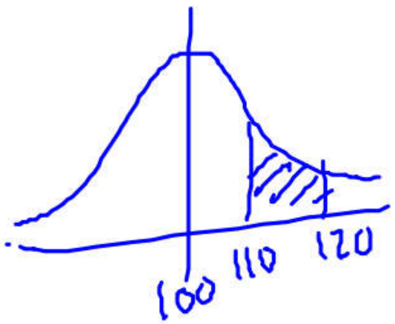
b) $P(X \geq 93) = P(X > 93)$

$\geq$ or $>$ doesn't matter

$= 0.6796$



c) $P(110 < X < 120) = 0.1613$

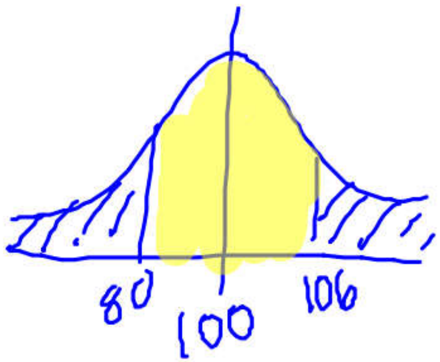


d) $P(X < 80 \text{ or } X > 106)$

$= 1 - P(80 < X < 106)$

$= 1 - 0.5642$

$= \underline{0.4358}$



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

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

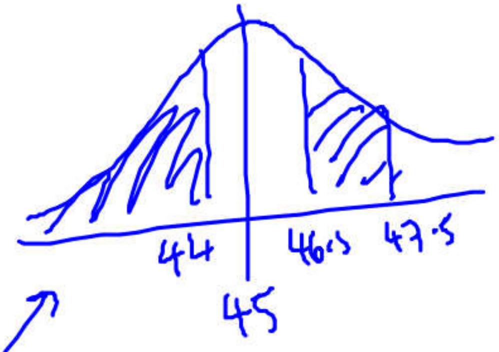
4 The random variable $X \sim \mathrm{N}(18, 10)$.

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

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

a Find $\mathrm{P}(T < 4.2)$.

b Without further calculation, write down $\mathrm{P}(T > 4.2)$.



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

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

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.

a) $P(X > 131) = 0.0194$

b) ~~[illegible]~~ We can use binomial distribution.

Y is the number of people who can join Mensa

$Y \sim B(30, 0.0194)$

$P(Y \geq 3) = 1 - P(Y \leq 2)$

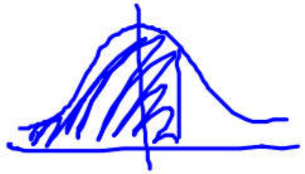
$= 1 - 0.9799$

$= 0.0201$

Ex 3B Q8-12

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?



DRAW A SKETCH!

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

a. $P(X < a) = 0.75$

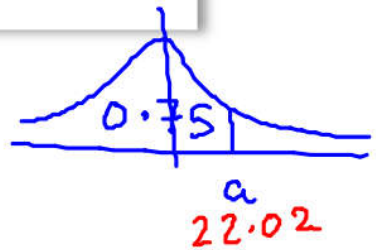
b. $P(X > a) = 0.4$

c. $P(16 < X < a) = 0.3$

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

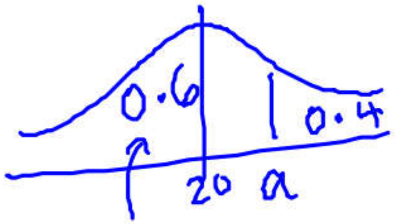
The graphics calculator is more advanced – you can assign whether the probability tails to the left (up to the value a), to the right (above the value a) or is symmetrically in the centre.

a) $P(X < a) = 0.75$



$a = 22.02$

b) $P(X > a) = 0.4$

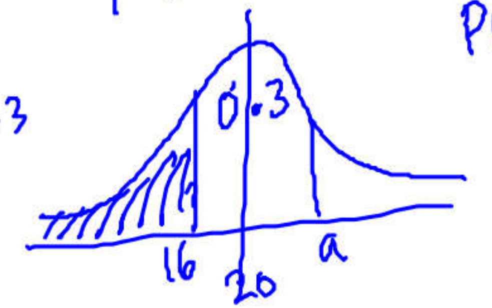


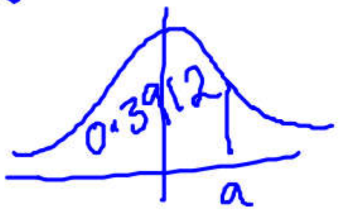
$a = 20.76$

$1 - 0.4$

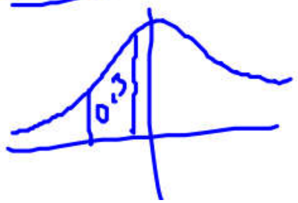
$P(X < 16) = 0.0912$

c) $P(16 < X < a) = 0.3$





$a = 19.17$

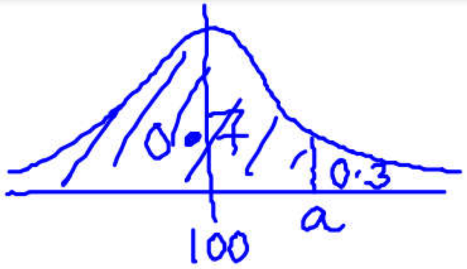


If the IQ of a population is distributed using $X \sim N(100, 15^2)$.

a. Determine the IQ corresponding to the top 30% of the population.

b. Determine the interquartile range of IQs.

a)



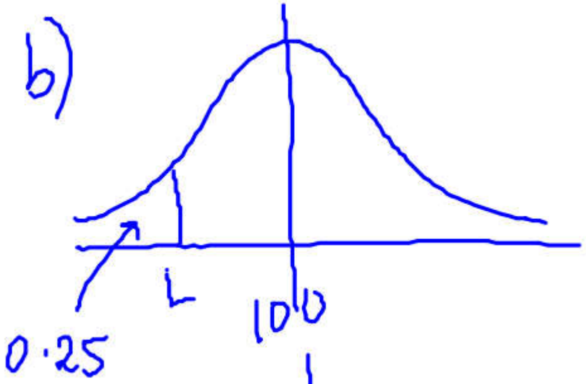
$P(X > a) = 0.3$

$P(X < a) = 1 - 0.3$

$= 0.7$

$a = 107.866 = \underline{\underline{108}}.$

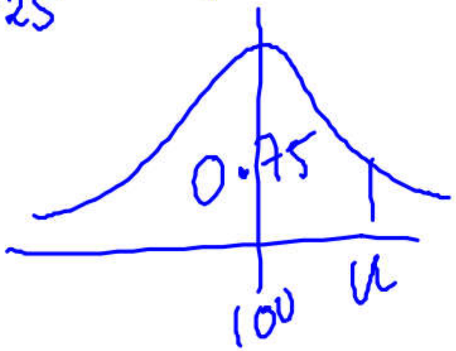
b)



$P(X < L) = 0.25$

$L = 89.88$

$\frac{2}{3} \times 15 = 10$



$P(X < u) = 0.75$

$u = 110.12$

$IQR = 110.12 - 89.88$

$= \underline{\underline{20.24}}$

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

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

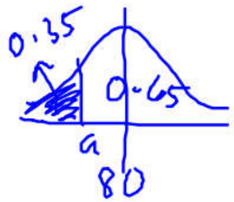
a. determine the a such that $P(X > a) = 0.65$

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.

a) $P(X > a) = 0.65$



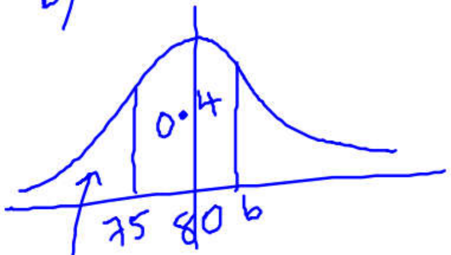
$P(X < a) = 0.35$

$\underline{\underline{a = 77.30}}$

b) $P(75 < X < b) = 0.4$

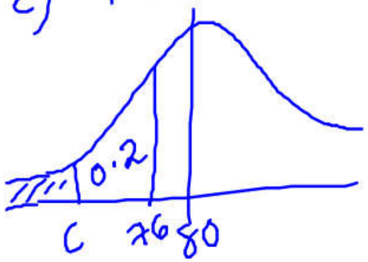
$P(X < b) = 0.4 + 0.2375 = 0.6375$

$\underline{\underline{b = 82.46}}$



$P(X < 75) = 0.2375$

c) $P(c < X < 76) = 0.2$



$P(X < 76) = 0.2839$

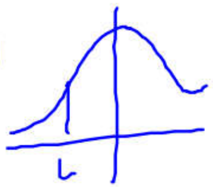
$P(X < c) = 0.2839 - 0.2$

$= 0.0839$

$c = 70.34$

Ex 3C

d)



$P(X < L) = 0.25$

$L = 75.29$

$P(X < U) = 0.75$

$U = 84.72$

$IQR = U - L = \underline{\underline{9.43}}$