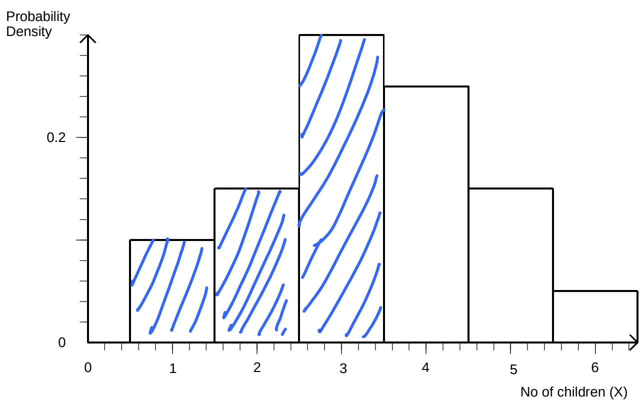
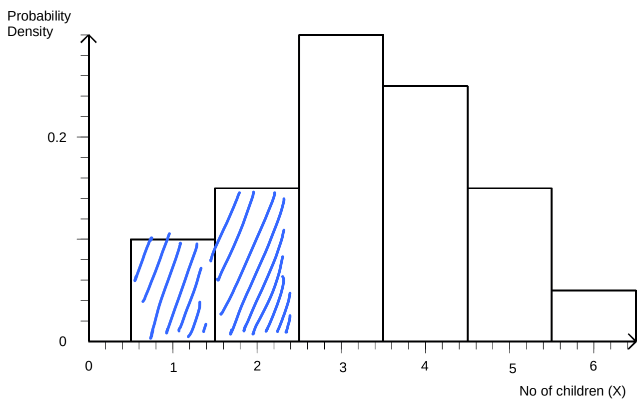
The Normal Distribution

Continuous Probability Distributions

A **discrete** probability distribution can be represented by a histogram, in which the AREA of each bar gives the probability for that value. So the total area of all the bars is 1. The scale on the vertical axis is **probability density** (like frequency density for a frequency histogram). Bars cannot be subdivided.

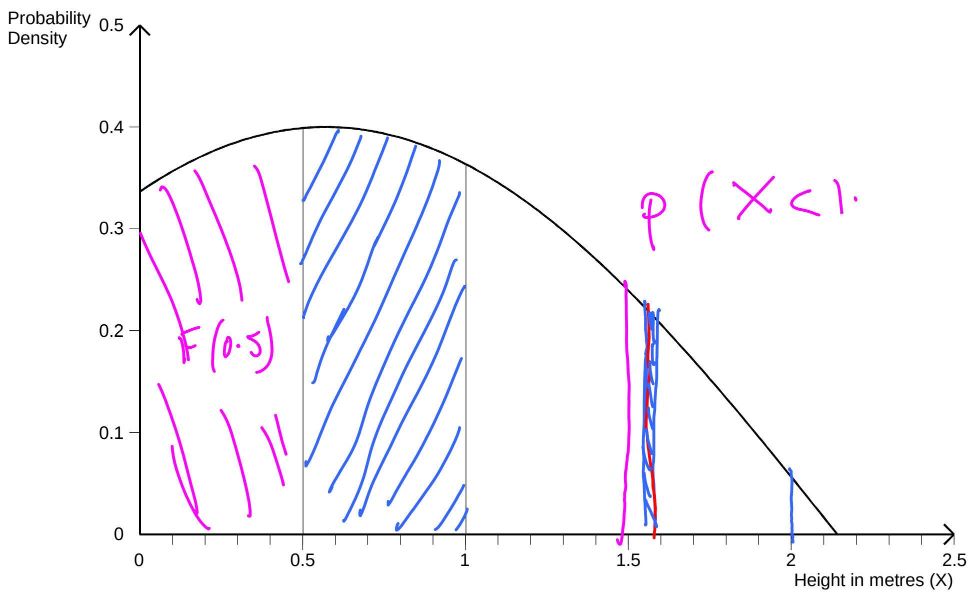


Area shaded is p(X ≤ 3)



Area shaded is p(X < 3)

A continuous probability distribution cannot be represented by bars since X can take an infinite number of different values. Instead it is represented by a probability density **curve**. Probabilities are represented by areas under the curve. So the total area under the curve is 1. We can draw vertical lines at any value of *x*, not just whole number values.



Area shaded is p(0.5 < X < 1)

It does not make sense to talk about p(X = 1.57) – this would be the area of a vertical line drawn at 1.57, which is zero.. When we say someone is 1.57m tall, we mean 1.565 < X < 1.575. We could calculate the probability of this. Also (unlike discrete distributions) there is no difference between p(X < 2) and p(X ≤ 2) since the only difference would be the area of a vertical line.

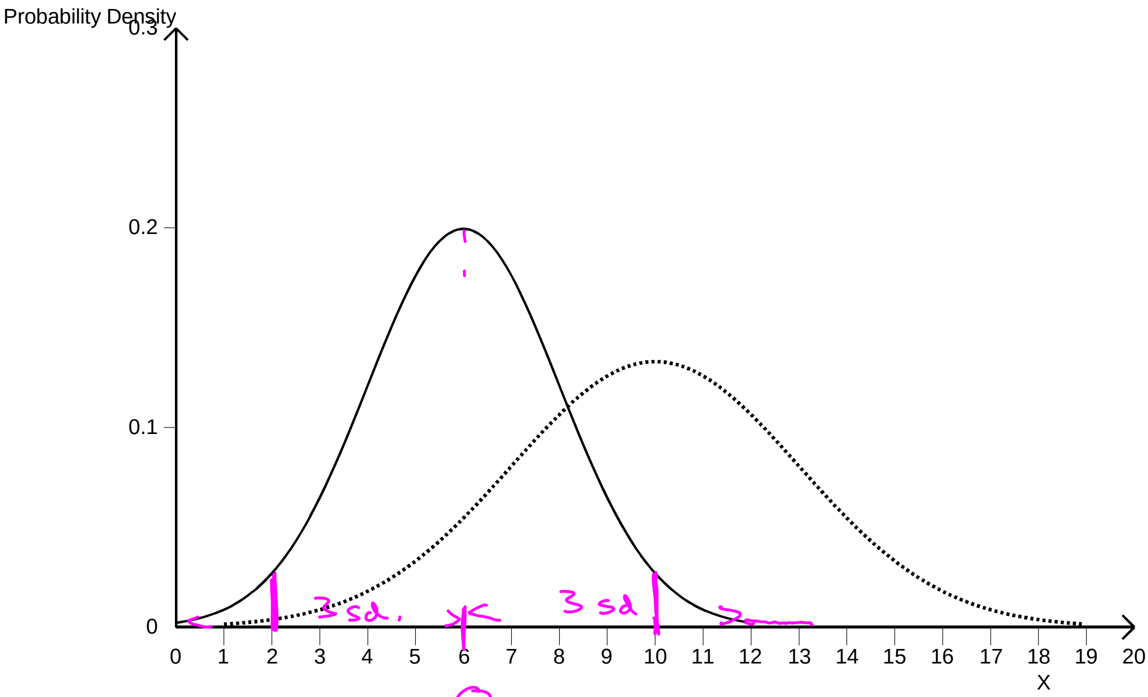
The equation of the curve is called the **probability density function** (pdf) of X, written f(x). The area to the left of a value x is called the **cumulative distribution function** (cdf) of X, written F(x). We can find the area between two limits by subtracting; eg the area above is F(1) – F(0.5).

In general we can find areas under the curve using integration. (in S2)

The Normal Distribution

This can be used to model many different real life variables. It has a bell-shaped curve, so that values are more likely to be near the mean than at the extremes. It has two parameters, the **mean μ** and **variance σ^2**. We write $X \sim N(\mu, \sigma^2)$

Changing these translates or stretches the curve without changing its basic shape.



Solid line: $X \sim N(6, 4)$ Dotted line: $X \sim N(10, 9)$

The curve is in theory asymptotic to the x-axis at each end. However, in practice, we can say that it is almost all contained within an interval of 3 standard deviations either side of the mean. For example, for $X \sim N(6, 4)$, the standard deviation is $\sqrt{4} = 2$, so the distribution is from $6 - 3 \times 2 = 0$ to $6 + 3 \times 2 = 12$. (To be precise, $p(0 < X < 12) = 0.9975$, so the area outside this interval is 0.0025)

Also 95% of the distribution lies within 2 standard deviations either side of the mean.

The pdf of the normal distribution is given the symbol $\varphi(x)$, and the cdf correspondingly $\Phi(x)$.

To find probabilities (ie areas), we use a table which gives areas under a "standard" Normal distribution $Z \sim N(0,1)$. The table gives $\Phi(z)$ (ie the area under the curve to the left of z) for different values of z.

To use this table with a different Normal distribution $X \sim N(\mu, \sigma^2)$, we have to convert X into Z using the transformation

$$z = \frac{x - \mu}{\sigma} \quad \text{or } x = \mu + z\sigma$$

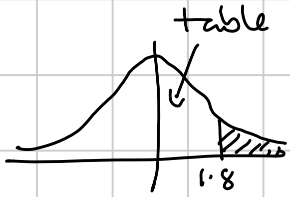
Examples

① If $Z \sim N(0,1)$, find

(a) $p(Z < 1.4) = 0.9192$

(b) $p(Z > 1.8) = 1 - 0.9641$

$= 0.0359.$



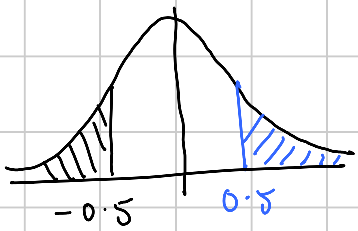
(c) $p(Z < -0.5)$

$= p(Z > 0.5)$

$= 1 - p(Z < 0.5)$

$= 1 - 0.6915$

$= 0.3085$

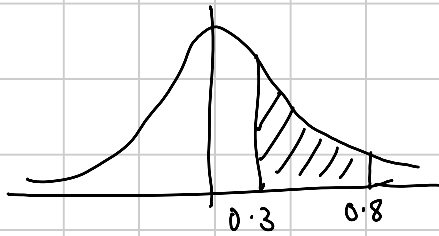


(d) $p(0.3 < Z < 0.8)$

$= \Phi(0.8) - \Phi(0.3)$

$= 0.7881 - 0.6179$

$= 0.1702$



$\boxed{\text{HWK} \quad \text{p 177 Ex 9A Q 1, 2}}$

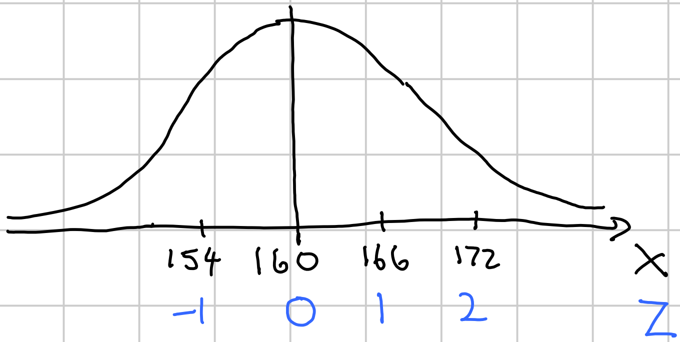
(2) The heights of 6^{th} form girls are Normally distributed with mean 160cm and variance 36cm^2. A girl is chosen at random. Find the probability that she is

(a) shorter than 166cm

Let X = height of girl

So $X \sim N(160, 36)$

$(\mu = 160, \ \sigma = 6)$



$$p(X < 166) = p\left(Z < \frac{166-160}{6}\right)$$

$$= p(Z < 1)$$

[Note Z measures "standard deviations above or below the mean"]

$$= 0.8413$$

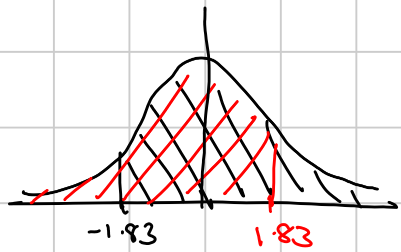
(b) taller than 149cm

$$p(X > 149) = p\left(Z > \frac{149-160}{6}\right)$$

$$= p(Z > -1.83)$$

$$= p(Z < 1.83)$$

$$= 0.9664$$

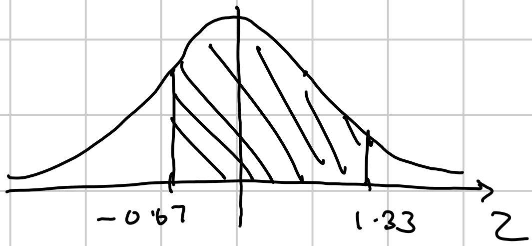


(c) between 156 and 168cm

$$p(156 < X < 168)$$

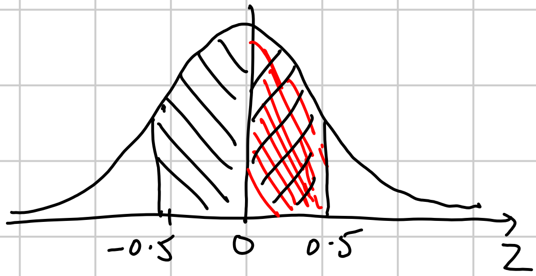
$$= p\left(\frac{156-160}{6} < Z < \frac{168-160}{6}\right)$$

$$= p(-0.67 < Z < 1.33)$$



$= p(Z < 1.33) - p(Z < -0.67)$
$= p(Z < 1.33) - p(Z > 0.67)$
$= p(Z < 1.33) - [1 - p(Z < 0.67)]$
$= 0.9082 - [1 - 0.7486]$
$= 0.6568$

(d) If there are 200 girls in the 6th form, how many would you expect to have heights within 3cm of the mean?

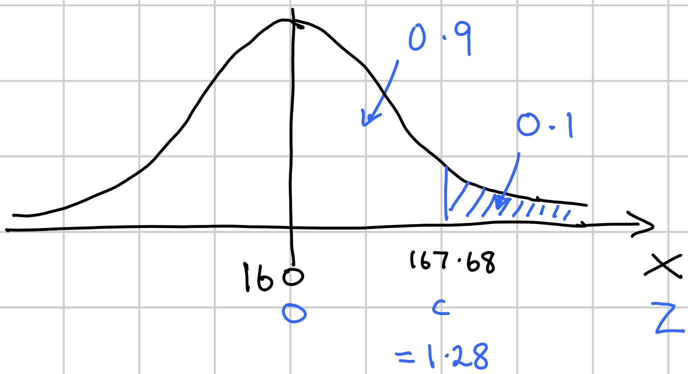


$p(157 < X < 163)$
$= p\left(\frac{157-160}{6} < Z < \frac{163-160}{6}\right)$
$= p(-0.5 < Z < 0.5)$

$= 2 \times p(0 < Z < 0.5)$
$= 2 \times [0.6915 - 0.5]$
$= 0.3830$

Expected no of girls $= 0.383 \times 200$
$= \underline{\underline{76.6 \text{ girls}}}$ (or 77 girls)

(e) Above what value would we expect the tallest 10% of girls' heights to lie?



We need $p(Z < c) = 0.9$
From the table,
$p(Z < 1.28) = 0.8997$

Therefore $\frac{X - 160}{6} = 1.28$
ie, $X = 160 + 1.28 \times 6$

$= \underline{\underline{167 \cdot 68 \text{ cm}}}$

NB We also have a small table which gives values of Z for certain common probabilities in the 'upper tail'. From this table we see that for an upper tail of 0.1, a more accurate value for Z is 1.2816. We should use these more accurate values if they are available.

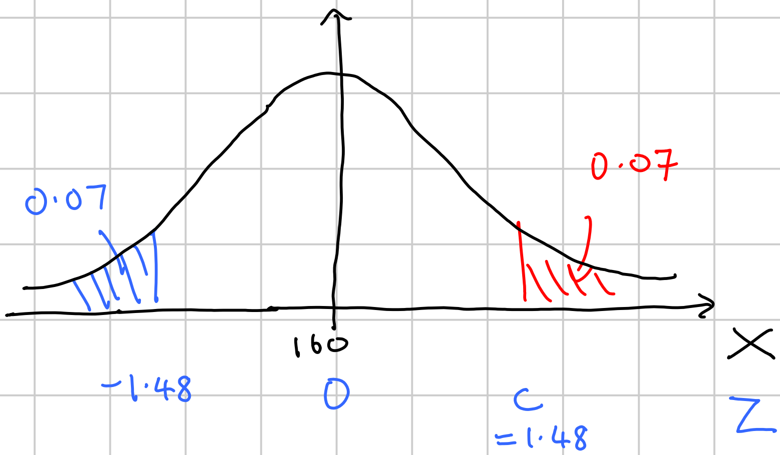
(f) Below what figure would we expect the shortest 7% of heights to lie?

look for $p(Z < c) = 0.93$

$\Rightarrow c = 1.48$

$\Rightarrow$ actual value required is -1.48

(ie 1.48 s.d's below the mean)

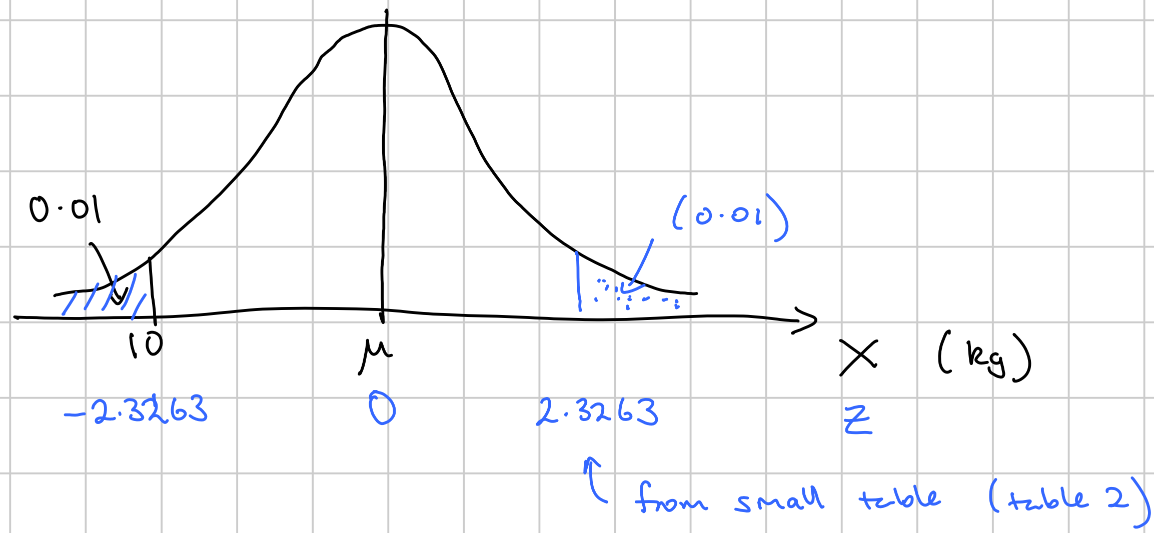


Required height is $X = 160 - 1.48 \times 6$

$= \underline{\underline{151 \cdot 12 \text{ cm}}}$

③ A firm sells bags of flour labelled as '10kg bags'. The amount of flour which the machine puts in each bag follows a Normal distribution with standard deviation 0·08 kg (ie 80g).

(a) Trading Standards Officers say that only 1% of bags sold can be underweight. What must be the mean amount of flour put into each bag?

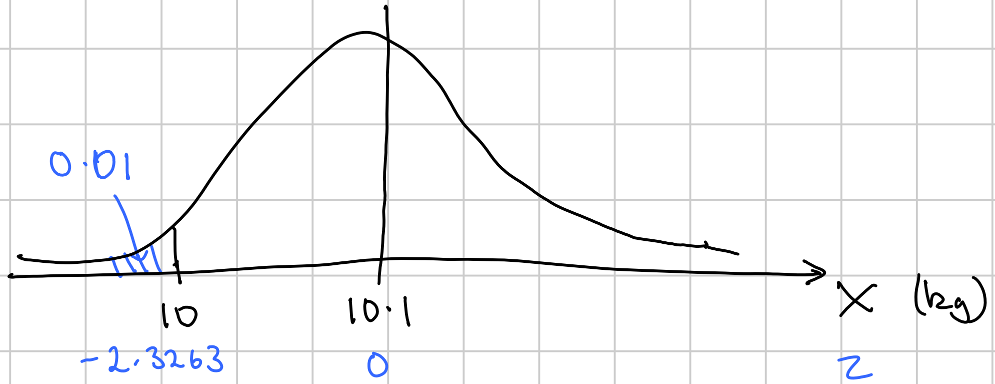


From diagram,

$$10 = \mu - (2 \cdot 3263 \times 0 \cdot 08)$$

$$\underline{\underline{\mu = 10 \cdot 2 \text{ kg}}} \ (3sf)$$

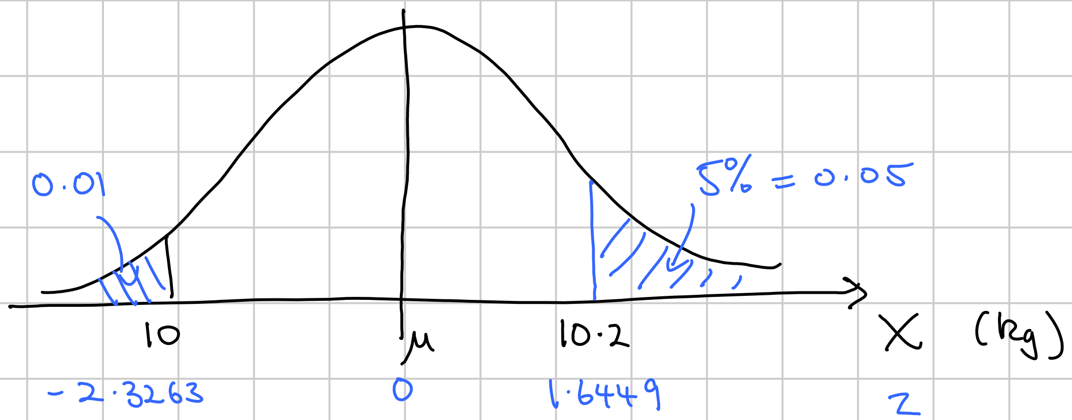
(b) The firm is unhappy that the mean is so high, so they decide to try to improve the accuracy of the machine so that they can set the mean to 10·1 kg. What should the new s.d be?



From diagram,

$$10 = 10.1 - 2.3263\sigma$$
$$2.3263\sigma = 0.1$$
$$\sigma = 0.0430 \quad (3\text{ s f})$$

(c) It proves impossible to make the machine this accurate, so as a compromise the company decides to set μ and σ so that 1% of bags are under 10 kg and 5% of bags are over 10.2 kg. What must μ and σ be to achieve this?



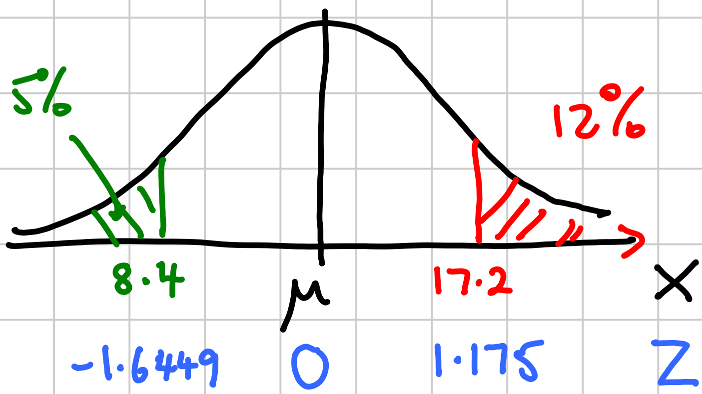
From diagram

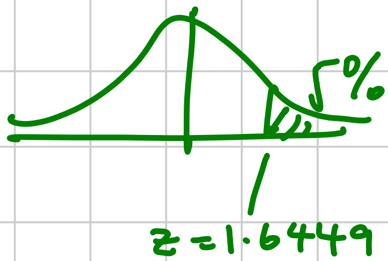
$$10 = \mu - 2.3263\sigma \quad ①$$
$$10.2 = \mu + 1.6449\sigma \quad ②$$

② − ①
$$0.2 = 3.9712\sigma$$
$$\underline{\underline{\sigma = 0.0504 \text{ kg.}}}$$
$$\underline{\underline{\mu = 10.12 \text{ kg}}}$$

p 178 Ex 9A Q 8 ab, 9 ab, 10, 16, 18

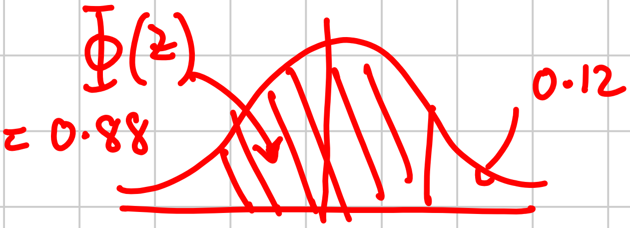
(4) A normal distribution is such that 5% of the values are below 8.4 and 12% of the values are above 17.2. Find μ and σ for this distribution.





For a 5% tail, look up $p = 0.05$ in the 'little table', and find $z = 1.6449$. Since this is a 'low end tail', mark on $z = -1.6449$

For a 12% tail, we cannot use the 'little table' so look in the 'big table' for $\Phi(z) = 0.88$



Since $\Phi(1.17) = 0.8790$
and $\Phi(1.18) = 0.8810$,
we can take $z = 1.175$

Form two equations:

$$17.2 = \mu + 1.175\sigma \quad (1)$$

$$8.4 = \mu - 1.6449\sigma \quad (2)$$

$(1) - (2)$ $\quad 8.8 = 2.8199\sigma$

$$\underline{\sigma = 3.12}$$

$$\underline{\mu = 13.5}$$

Ex 9A Q 10, 11, 12 18
by next Wednesday.