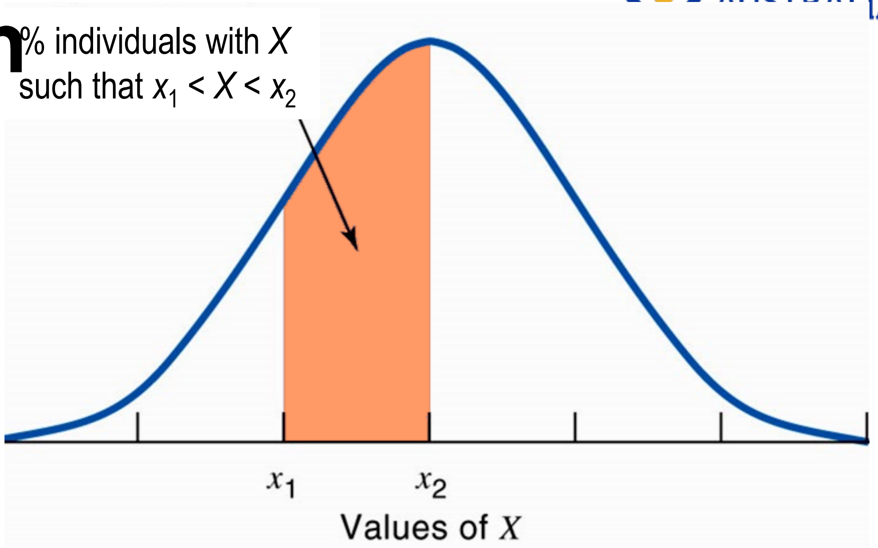
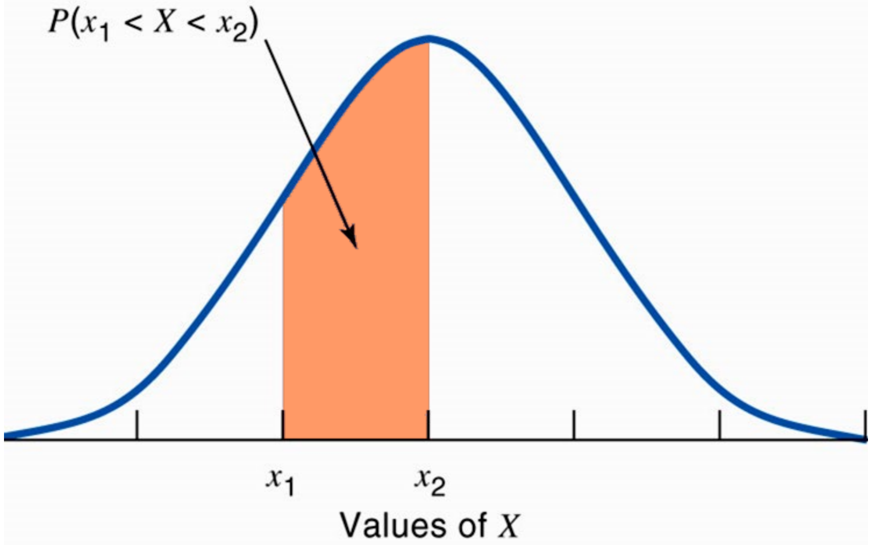
Continuous random variable and population distribution

The shaded area under a density curve shows the proportion, or %, of individuals in a population with values of X between x_1 and x_2.



Because the **probability** of drawing **one individual at random** depends on the **frequency** of this type of individual in the population, the probability is also the shaded area under the curve.

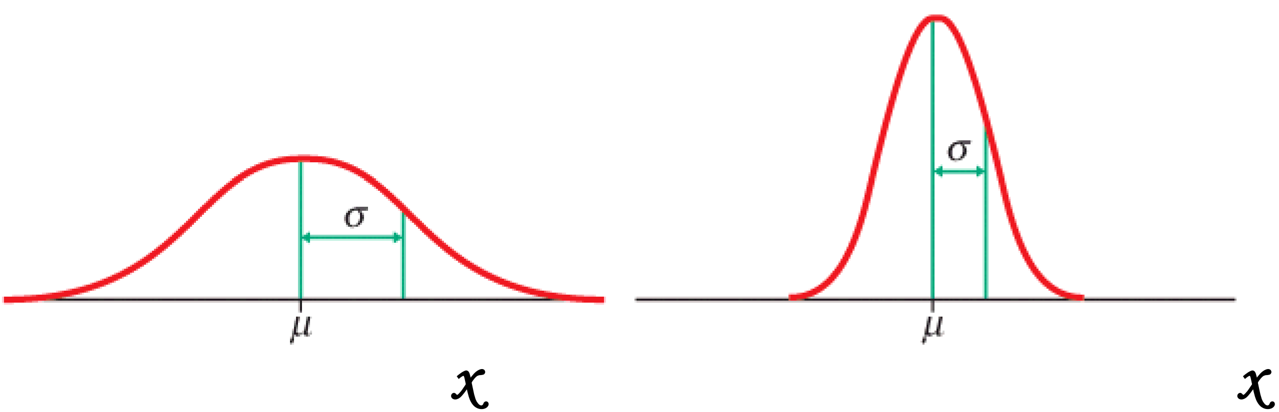


Normal distributions

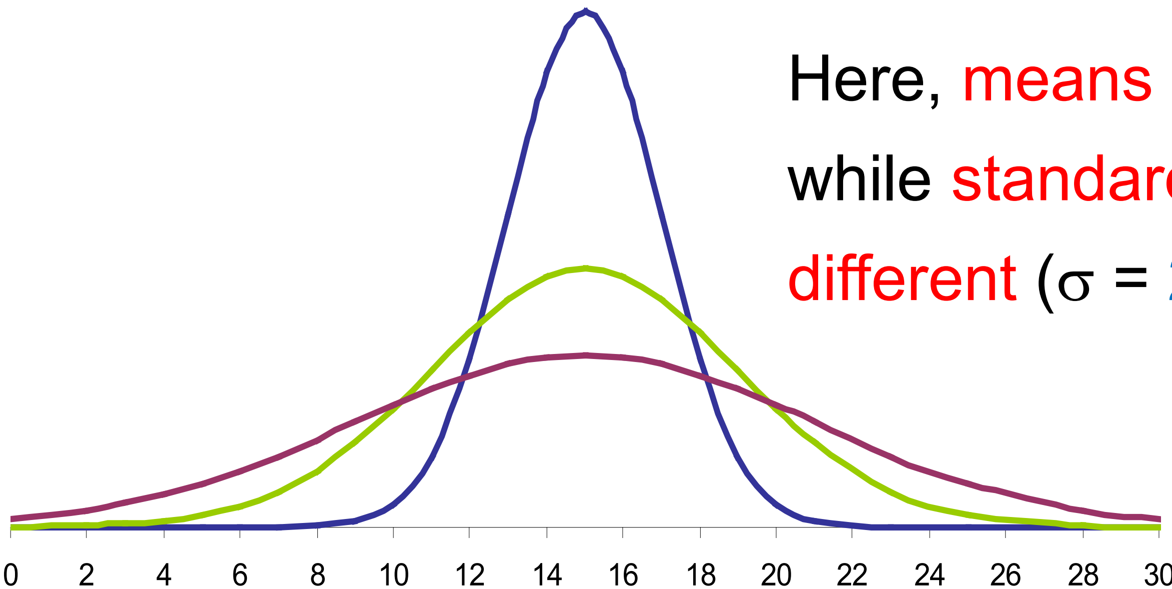
- Normal distributions are **symmetrical, bell-shaped** density curves defined by
  - Center: a mean μ (*mu*) and
  - Spread: a standard deviation σ (*sigma*)

A random variable X is distributed as $N(\mu,\sigma)$:

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

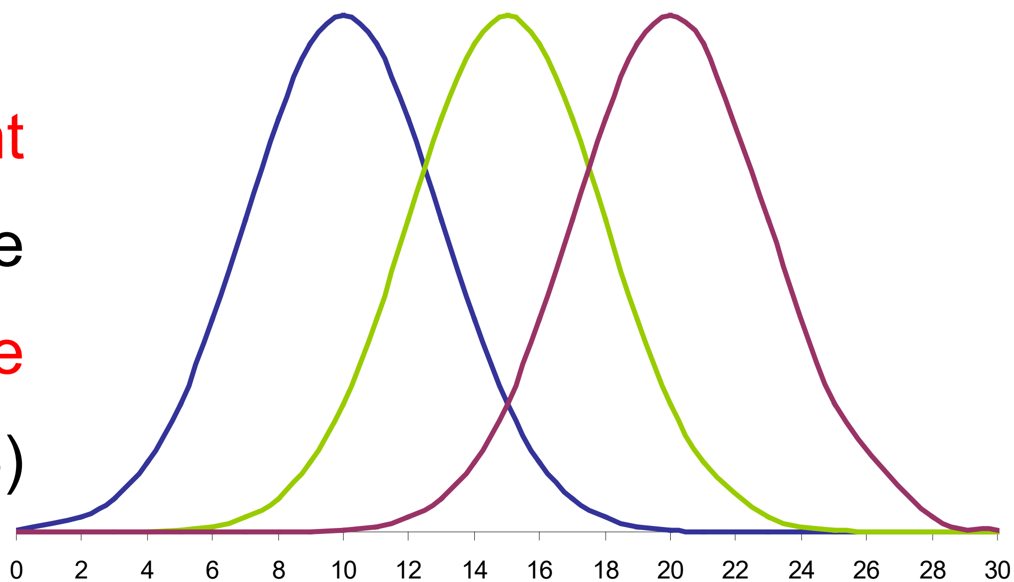


A family of Normal density



Here, means are the same (μ = 15) while standard deviations are different (σ = 2, 4, and 6).

Here, means are different (μ = 10, 15, and 20) while standard deviations are the same (σ = 3)



The Normal Distribution probabilities

- Area under sections of the Normal curve is the probability
  [i.e. proportion (or %) of subjects in a population]

- How to calculate these Normal probabilities?
  - dnorm PDF
  - pnorm CDF
  - qnorm Quantile
  - rnorm Random Normal

The Normal Distribution in R

Description

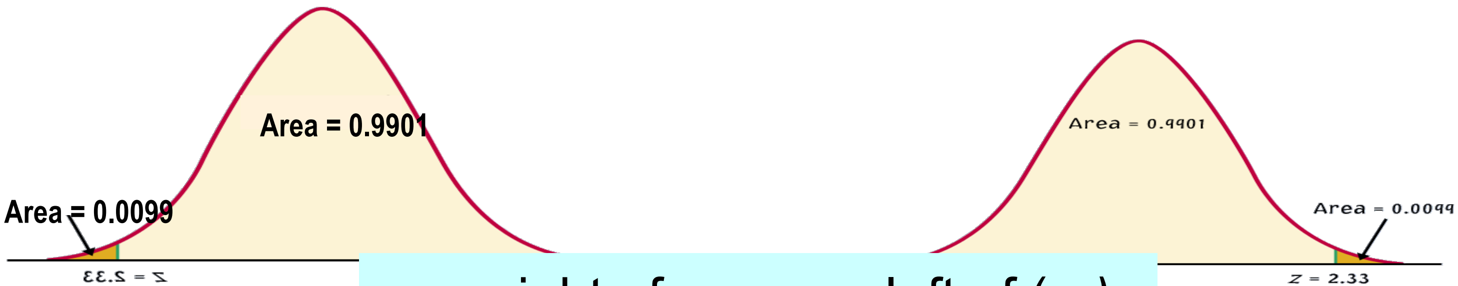
Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.

Usage

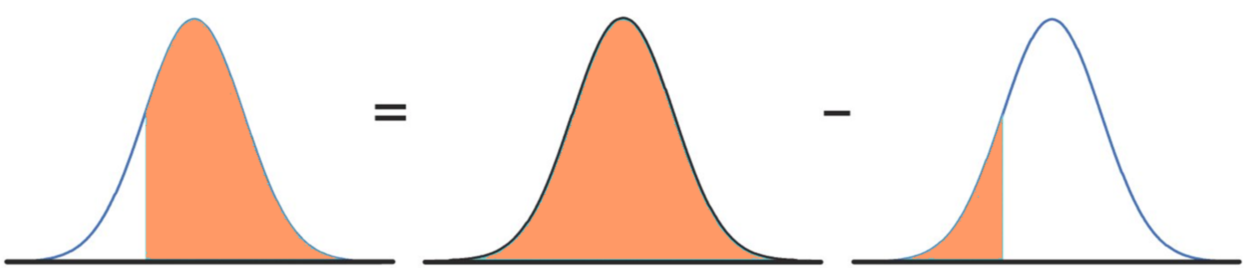
```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Tips on Calculating Normal Probability

Because the Normal distribution is symmetrical, there are 2 ways that you can calculate the area under the standard Normal curve <u>to the right</u> of a z value.



area right of z = area left of $(-z)$

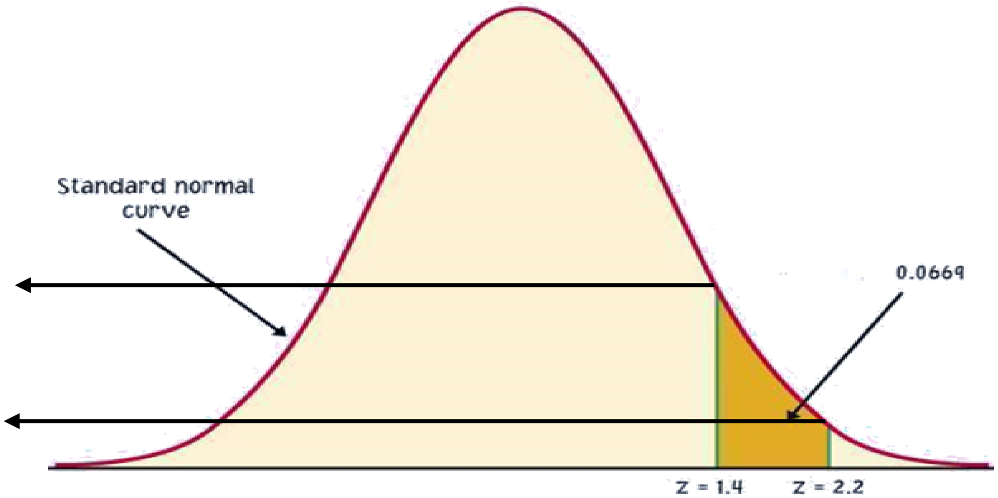


area right = 1 - area left

Tips on Calculating Normal Probability

To calculate the area between two z- values, first get the area under N(mean,sd) the left for each z-value from R.

Then subtract the smaller area from the larger area.



area between z_1 and z_2 =
area left of z_2 – area left of z_1

In R: **pnorm(z2, mean, sd) - pnorm(z1, mean,sd)**

Example 1

A random variable, Y, denoting Year 3 test scores, has a **Normal distribution with a mean of 70 and a standard deviation of 10**.

The probability of a Year 3 child scoring **below 50** is closest to:

A. 2.0

B. -2.0

C. 0.95

D. 0.05

E. 0.023

```
> answer=pnorm(50, 70,10)
[1] 0.02275013
```

Example 2

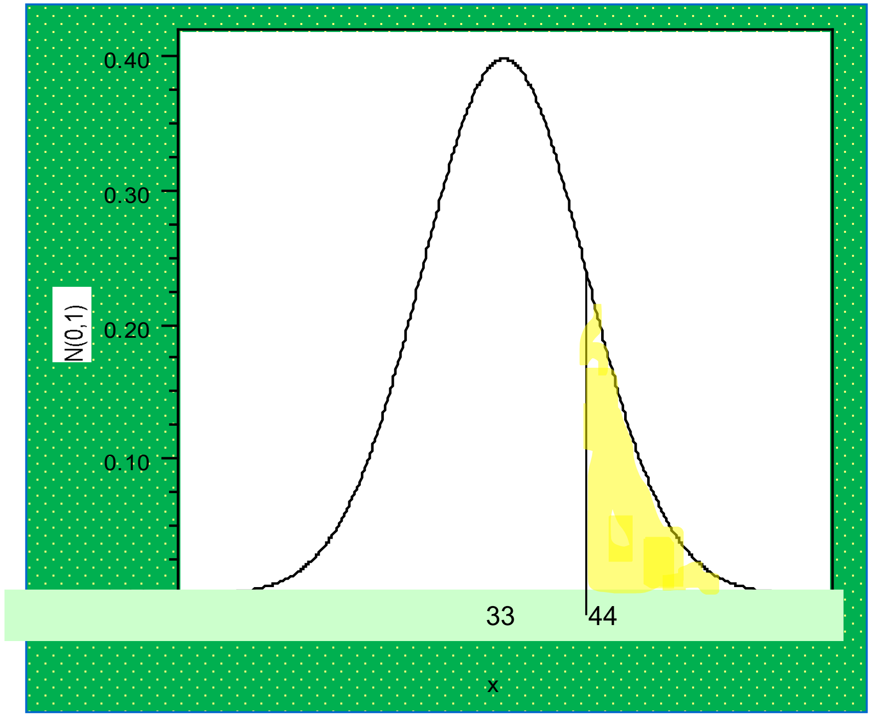
Suppose we know that the population of fish has a mean length *(μ)* of 33cm with a standard deviation *(σ)* of 11cm.

Thus $X \sim N(\mu=33\text{cm}, \sigma=11\text{cm})$

If we randomly selected a fish from this population, what is the probability of getting a fish longer than 44cm?

Normal (mean=33.sd=11)

$$X \sim N(\mu = 33, \sigma = 11)$$



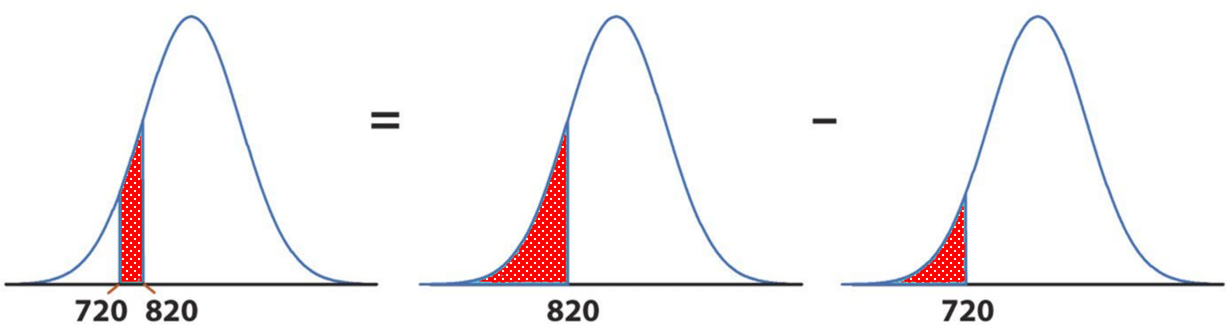
answer=1- pnorm(44, 33,11)

= 1 - 0.8413447 =0.1586553

Example 3. The National Collegiate Athletic Association (NCAA) requires Division I athletes to score at least 820 on the combined math and verbal SAT exam to compete in their first college year. The SAT scores of 2003 were approximately Normal with mean 1026 and standard deviation 209.

The NCAA defines a "partial qualifier" eligible to practise and receive an athletic scholarship, but not to compete, with a combined SAT score of at least 720.

What proportion of all students who take the SAT would be partial qualifiers? That is, what proportion have scores between 720 and 820?



area between 720 and 820 = area left of 820 - area left of 720

= 0.1622 - 0.0716

≈ 9%

About 9% of all students who take the SAT have scores between 720 and 820.

Detail

- X is Normal (mean=1026, sd=209)
- $P(720 < X < 820)$

= pnorm(820, 1026, 209) - pnorm(720, 1026, 209)

= 0.1621534 - 0.07158129

= 0.09057216

```
> pnorm(820, 1026, 209)
[1] 0.1621534

> pnorm(720, 1026, 209)
[1] 0.07158129
```

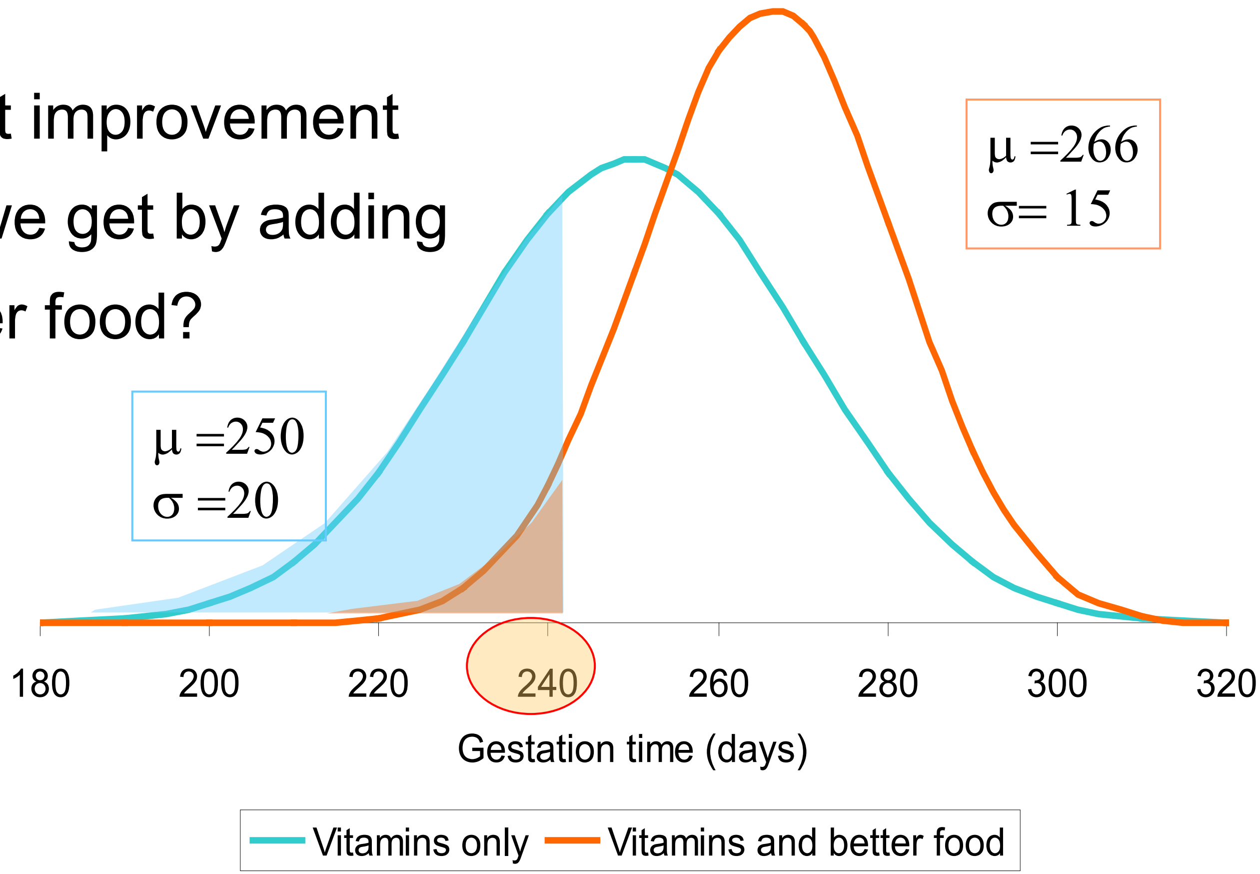
Example 4. Gestation time in malnourished mothers

THE UNIVERSITY OF WESTERN AUSTRALIA

What is the effect of better maternal care on gestation time and preemies?

The goal is to obtain pregnancies 240 days (8 months) or longer.

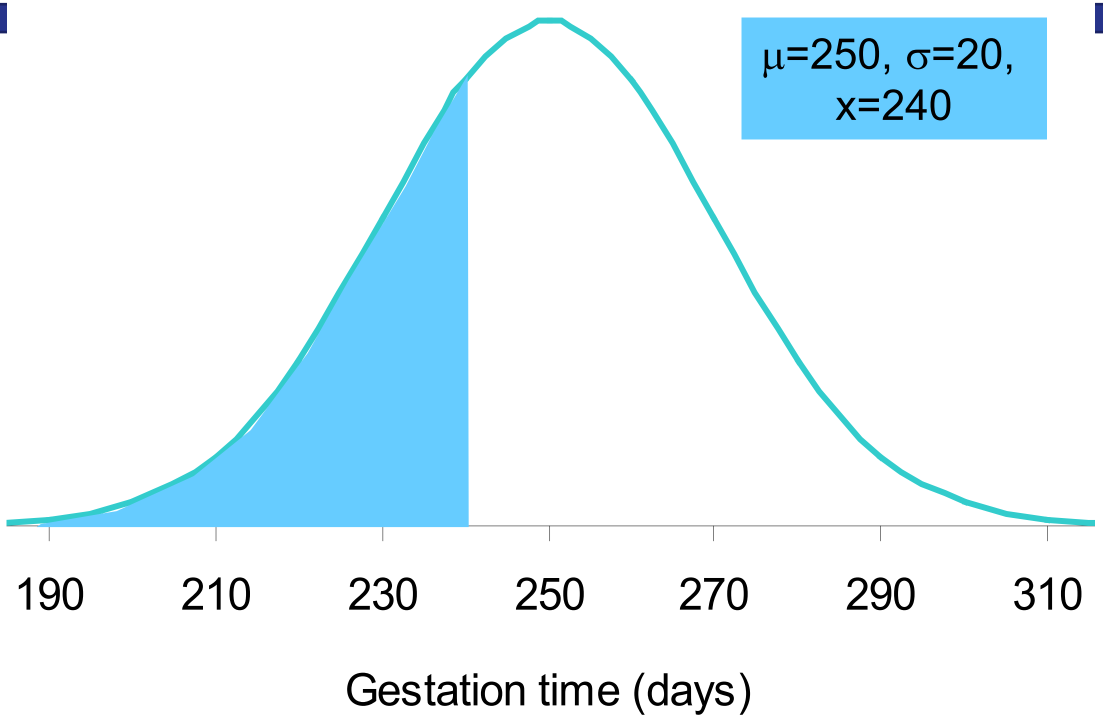
What improvement did we get by adding better food?



(a) Under **each treatment**, what percentage of mothers failed to carry their babies at least 240 days?

Vitamins Only

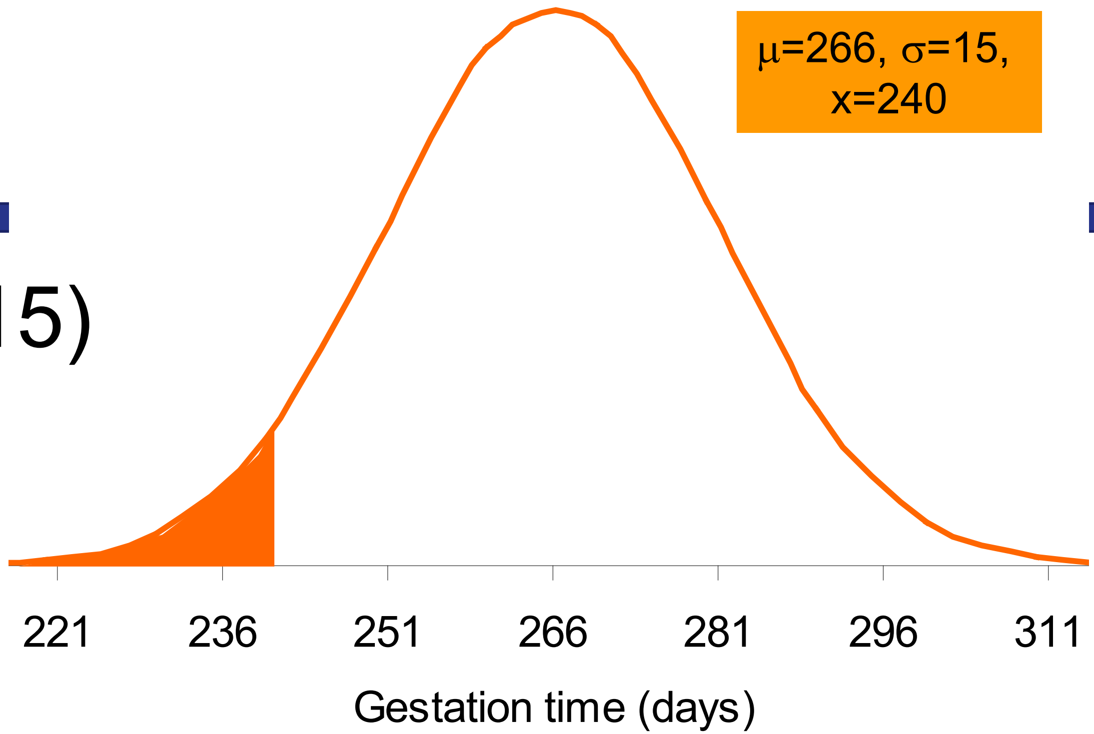
```
> pnorm(240,250,20)
[1] 0.3085375
```



<u>Vitamins only</u>: 30.85% of women would be expected to have gestation times shorter than 240 days.

(b)Vitamins and better food

```
> pnorm(240,266,15)
[1] 0.04151822
```



Vitamins and better food: 4.15% of women would be expected to have gestation times shorter than 240 days.

Compared to vitamin supplements alone, vitamins and better food resulted in a much smaller percentage of women with pregnancy terms below 8 months (4% vs. 31%).

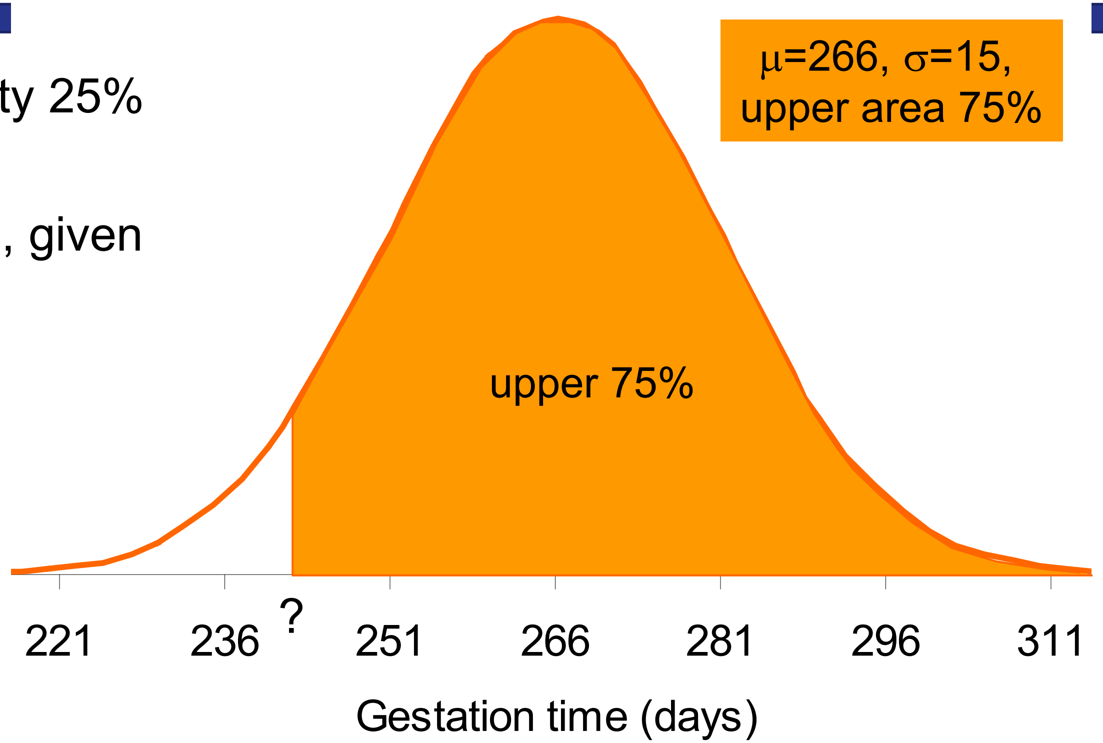
Vitamins and better food

Example 9 Continued (c) How long are the longest 75% of pregnancies when mothers with malnutrition are given vitamins and better food?

We are provided with the probability 25% (cumulative)

We are looking for the x=gestation, given 25%

Use qnorm




```
> qnorm(0.25,266,15)
[1] 255.8827
```

256 days

➔ The 75% longest pregnancies in this group are 256 days or longer.