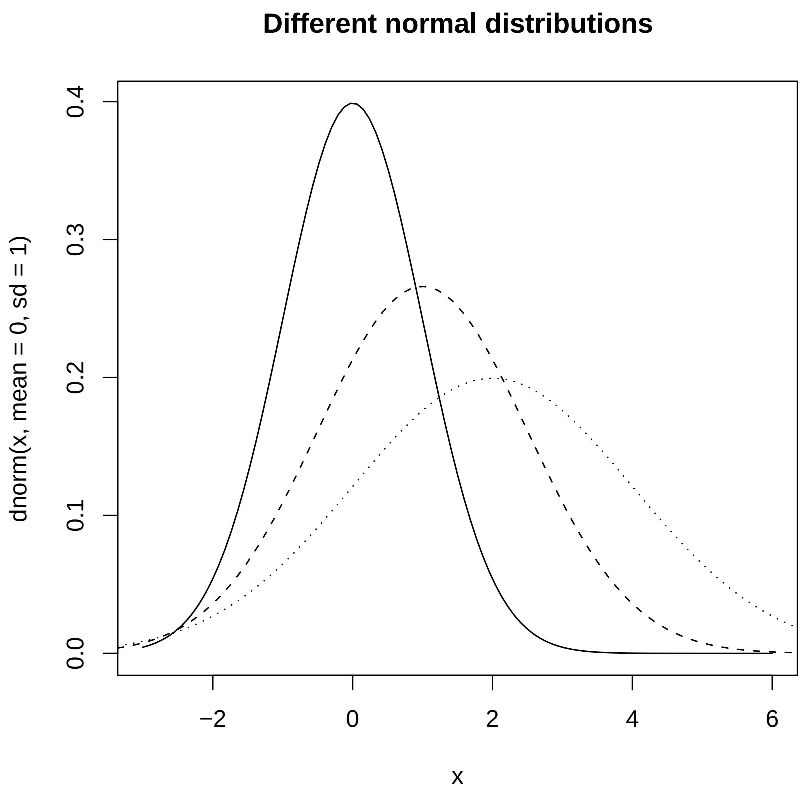
The normal distribution describes *normally distributed numbers*. A normal distribution is described by two **parameters**, the mean μ and standard deviation σ. In general, the density is a *bell-shaped curve*, centered at μ with a standard deviation of σ.

The picture below shows a few normal densities with different μ and σ values.

```
> curve(dnorm(x, mean = 0, sd = 1), -3, 6)
> curve(dnorm(x, mean = 1, sd = 1.5), add = TRUE, lty = 2)
> curve(dnorm(x, mean = 2, sd = 2), add = TRUE, lty = 3)
> title("Different normal distributions")
```



0.1 Problems

1. Which density in the figure is for the distribution with mean of 2?
2. Which density in the figure is for the distribution with the smallest standard deviation?

1 Probability

The density answers questions about probabilities involving a continuous random variable X according to

$$P(a \leq X \leq b) = \text{ The area under } f(x) \text{ from } a \text{ to } b.$$

In class we learned how to use a table to answer questions about areas. We'll see it is easier with the computer.

Let X represent the gestation time for an elephant. Assume this has a normal distribution with mean 641 days and a 14 days standard deviation. What is the probability that a gestation period will be less than 630 days? Between 630 and 650 days? More than 650 days?

In R, such questions are answered using the `pnorm()` function. Unlike the book, this answers $P(X \leq x)$ and not $P(0 \leq X \leq x)$. So that in particular

$$P(a \leq X \leq b) = \texttt{pnorm(b,mean,sd) - pnorm(a,mean,sd)}$$

So for the three questions above we have

```
> pnorm(630, mean = 641, sd = 14)

[1] 0.2160174

> pnorm(650, mean = 641, sd = 14) - pnorm(630, mean = 641, sd = 14)

[1] 0.5238242

> 1 - pnorm(650, mean = 641, sd = 14)

[1] 0.2601584
```

2 Problems

1. For the same elephant, find the probability that a gestation time is less than 615 days.

2. Find the probability that the gestation time for the elephant is between 613 and 669 days.

3. Normal body temperature varies from person to person. Suppose a randomly chosen *healthy* adult will have a body temperature that is normally distributed with mean 98.2 and standard deviation 0.7. Find the probability

   (a) That a randomly chosen healthy adult will have a temperature more than 100 degrees?

   (b) That a randomly chosen healthy adult will have a temperature more than 99 degrees.

2.1 quantiles

The reverse problem – find the value splitting the top x% from the bottom 100-x% – is found using quantiles, and the `qnorm()` function. For instance, if we want to know the gestation length that 75% of Elephants deliver by we have

```
> qnorm(0.75, mean = 641, sd = 14)
```

```
[1] 650.4429
```

Where as the top 10% of lengths is found with

```
> qnorm(1 - 0.1, mean = 641, sd = 14)
```

```
[1] 658.9417
```

2.2 Problems

1. For the normal distribution with mean 0 and standard deviation 1 the IQR covers the middle 50% of the data. Find the value of the IQR. (Hint finding the 75th percentile, or 0.75 quantile will help answer this.)

2. If heights are normally distributed with a mean of 70 and standard deviation of 3, find the height corresponding to the 10th percentile (10 percent shorter, 90 percent taller).

3 Random samples

It is easy to get lost with densities and areas as to what they say. They are describing random numbers. The `rnorm()` function can be used to randomly sample random numbers from a normal distribution.

For instance

```
> rnorm(1, mean = 631, sd = 14)
```

```
[1] 608.8765
```

Is a randomly chosen number from the elephant-gestation distribution. If you run this command, you will certainly get a different number, as these are *random* numbers.

How probability comes in is when you try to describe how big that number will be. In this case, roughly 95% of the time the number will have a z-score between -2 and 2, so between $641 - 2 * 14$ and $641 + 2 * 14$. (Why?)

To randomly sample lots of numbers, we just ask for that many. For instance, this samples 50 such numbers and then computes how many are less than 1 standard deviation. (We expect to get 68% plus 16%, or 84%. Why?)

```
> x = rnorm(50, mean = 641, sd = 14)
> sum(x < 641 + 14)

[1] 42

> sum(x < 641 + 14)/50 * 100

[1] 84
```

The density plot is a good way to view random samples. If you are using pmg, drag the variable `x` to the Lattice Explorer and make a density plot. The points are plotted on the bottom. They should be clustered near the mean, then thin out as there is more area around the mean, than not. The thinning is determined by the width of the graph, which is the standard deviation.

4 Problems

1. Repeat this command

   ```
   > rnorm(1)
   ```

   Until you get a value more than 1.5. How many times did it take? What is the probability that it happened the first time?

2. Generate 50 random samples of your own, as above. Find the percentage that are less than 1 standard deviation above the mean. Compare your answer to 84%.

3. Generate 1000 random samples of your own, modifying the above commands. Find the percentage that are less than 2 standard deviations above the mean. What do you expect this percent to be? Does it compare