**2. Statistical Distributions**

**AS – Discrete Distributions**

**The Binomial Distribution**

For a fixed number of trials, *n*, each with a probability *p* of occurring, the probability of a number *x* of successes is given by the formula

Binomial distribution tables (and calculators) give you cumulative probability P(X ≤ x)



Eg1 The random variable X ~ B(10, 0.35), find:

1. P(X ≤ 6)
2. P(X ≥ 5)
3. P(X = 6)
4. P(4 ≤ X ≤ 7)

The binomial distribution can be appropriately applied under the following conditions:

* the trials are independent
* the trials have a constant probability of success
* there are a fixed number of trials
* there is only success or failure.

**The Poisson Distribution**

The Poisson distribution is a discrete probability distribution which is used to model the number of events occurring randomly within a given interval of time and space.

In a particular interval, the probability of an event *X* occurring *x* number of times is given by:

where λ = μ = E(X) and x = 0, 1, 2, 3, …

If the probabilities are distributed in this way, it is written X ~ Po(λ)

As with the binomial distribution, tables give you the cumulative probability P( X ≤ x)

Eg2 The number of telephone calls received at an exchange during a weekday morning follows a Poisson distribution with a mean of 6 calls per 5 minute period. Find the probability that

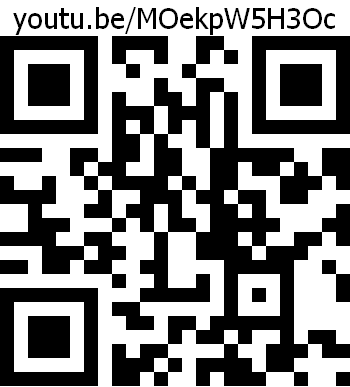
1. there are no calls received in the next five minutes
2. 3 calls are received in the next five minutes
3. fewer than 2 calls are received between 11:00 and 11:05
4. more than 2 calls are received between 11:30 and 11:35

The Poisson distribution can be appropriately used when

* *n* is large (usually > 50) **and**
* *p* is small (usually < 0.1)

**The Poisson distribution can be used as an approximation to the binomial distribution**

If X ~ B(*n, p*) with large *n* and small *p*, then **X ~ Po(*np*)**

Eg3 The probability that a wrapped chocolate biscuit is double wrapped is 0.01. Use a suitable approximation to find the probability that of the next 60 biscuits you unwrap:

1. none are double wrapped
2. at least 2 are double wrapped

**The Discrete Uniform Distribution**

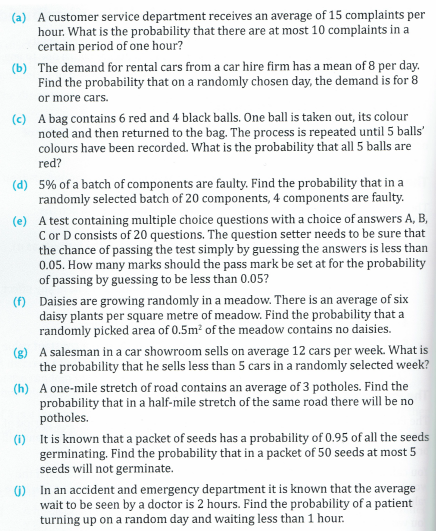
A discrete uniform distribution is a distribution where all the outcomes are equally likely, for example the outcome when a fair dice is thrown.

If X is a discrete variable and is uniformly distributed on the set {1, 2, 3, 4, …, N} then

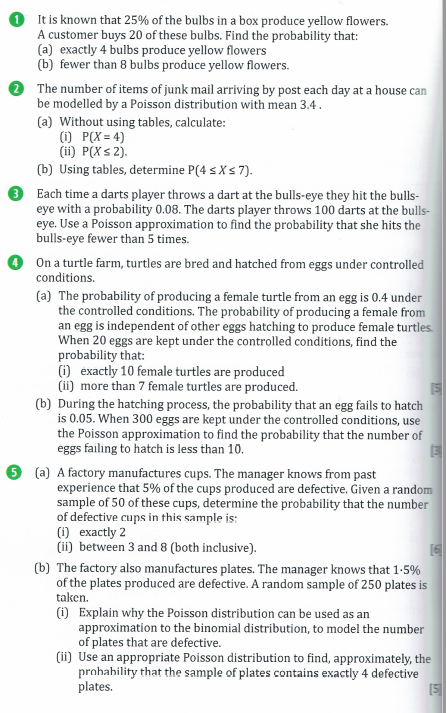
Eg4 A fair octagonal spinner numbered from 1 to 8 is spun and the number obtained X is recorded. This process is repeated a set number of times.

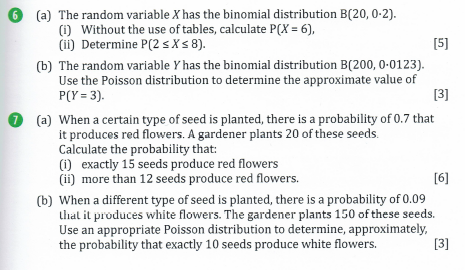
Find P(2 ≤ X < 5)

At both AS and A2 you may be given questions where the appropriate distribution to apply is not provided. In each of the following situations decide which are best modelled by binomial, Poisson or uniform distributions. Give a reason for your decision.



Exercise 2.1





Numerical Answers

(1a) 0.1896 (b) 0.8982

(2ai) 0.1858 (ii) 0.3397 (b) 0.4185

(3) 0.0996

(4ai) 0.1171 (ii) 0.5841 (b) 0.0699

(5ai) 0.2611 (ii) 0.4587 (b) 0.194

(6ai) 0.109 (ii) 0.9208 (b) 0.212

(7ai) 0.179 (ii) 0.772 (b) 0.076

**A2 – Continuous Distributions**

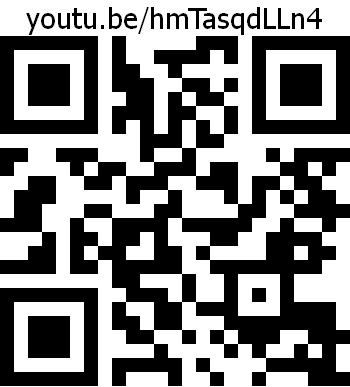
**The continuous uniform (rectangular) distribution X ~ U[a, b]**

This has a constant probability density function (pdf) over a range of values and zero elsewhere.

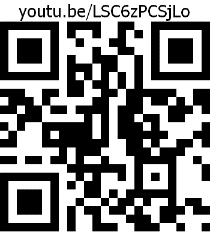


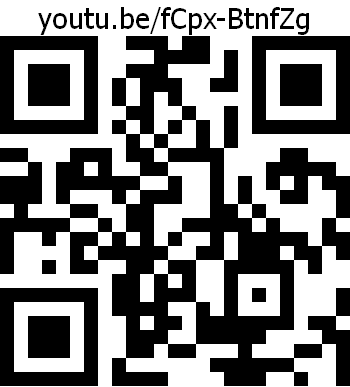
Eg5 The continuous variable X is uniformly distributed X ~ U[2, 5]

Find (a) E(X) (b) Var(X) (c) P(X > 3.8)

Eg6 A junior gymnastics league is open to children who are at least 5 years old but have not yet had their 9th birthdays. The age X years, of a member is modelled as a uniform continuous distribution over the range of possible values between five and nine. Age is measured in years and decimal parts of a year, rather than just completed years. Find

1. the pdf f(x) of X
2. P(6 ≤ X ≤ 7)
3. E(X)
4. Var(X)
5. The percentage of the children whose ages are within one standard deviation of the mean.

Eg7 A piece of string of length 8cm is randomly cut into two pieces. Find the probability that the longer of the two pieces of string is at least 5cm long.



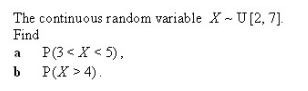
Eg8 Given that Y ~U[a, b] and E(Y) = 3 and Var(Y) = 3, find P(Y < 2).

Eg9 The amount of time, in minutes, that a person must wait for a bus is represented by the pdf

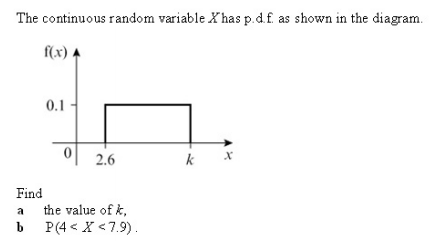
T ~ U[0, 15].

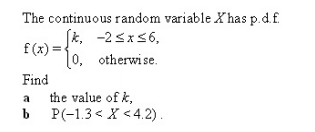
1. what is the probability that the person waits fewer than 12.5 minutes?
2. On average, how long must a person wait.
3. What is the standard deviation of the waiting time?
4. 90% of the time, the time a person must wait falls below what value?

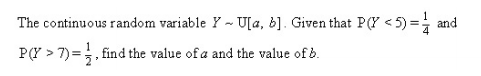
**Exercise 2.2**



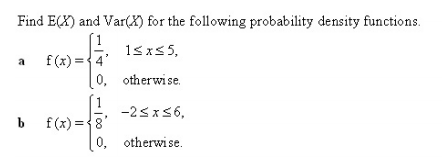
1.

2. 

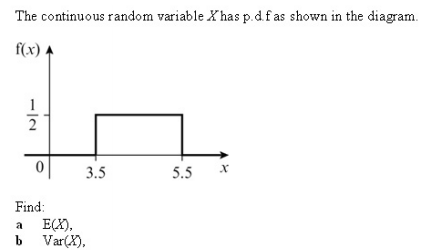
3. 



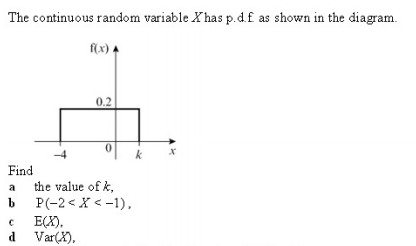
4.



5.

6.

7. 



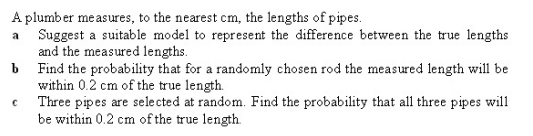
Numerical Answers

(1a) 0.4 (b) 0.6 (2a) 12.6 (b) 0.39 (3a) 1/8 (b) 0.6875 (4) a = 3, b = 11

(5a) 3, 4/3 (b) 2, 16/3 (6a) 4.5 (b) 1/3 (7) a = -1, b = 3

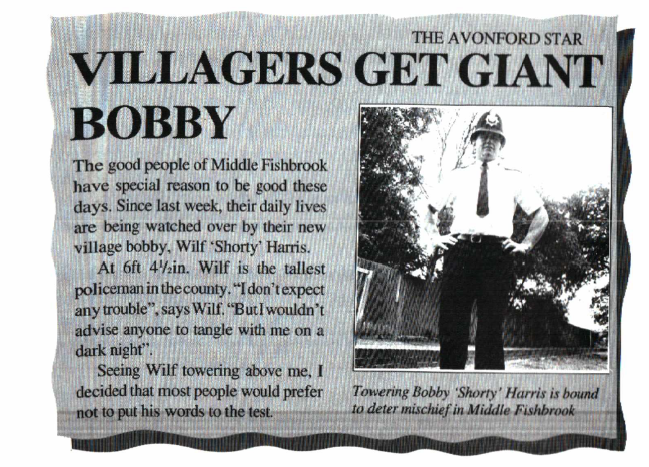
(8a) k = 1 (b) 0.2 (c) -1.5 (d) 25/12 (9b) 0.4 (c) 0.064

8.

9. 

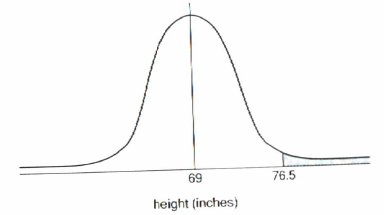
**The Normal Distribution**

Just as the binomial and Poisson distributions are important examples of the special distributions of the discrete kind, so the Normal distribution can be described as the single, most important continuous distribution in statistics. The form of the data approximates very well to data of the ‘natural phenomenon’ type, such as weights, heights and ages; data that occurs naturally in all types of situations.



Wilf is clearly very tall, but how much so? Is he one in a hundred, or a thousand or even a million?

To answer this question we need to know the distribution of the heights of adult British men. This may be modelled by the Normal distribution which has the distinctive pdf shown below:



As always with probability density functions, the area beneath the curve represents probability, so the shaded area to the right of 76.5in represents the probability that a randomly selected adult male is over 6ft 4½ inches tall.

Before we are able to find this area we need to know the mean and standard deviation of the distribution. For adult British males these are 69 inches and 2.5 inches respectively.

We can summarise this as

for the continuous random variable H, where H ~ N(69, 2.52), find P(H > 76.5)

we can use our calculators:

**Normal CD**

**Lower = 76.5**

**Upper = 1 x 1099**

**σ = 2.5**

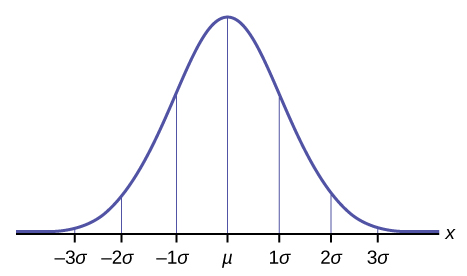
**μ = 69**

which will produce P(H > 76.5) =

So the probability of a randomly selected adult male being at least as tall is Wilf is 0.0013, ie just over 1 man in a thousand.

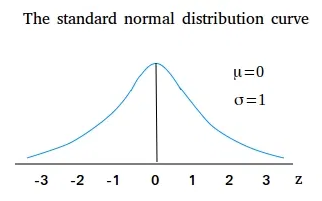
The key properties of a Normal distribution can be summarised as:

* The distribution is symmetrical about the mean, μ
* The mode, median and mean are all equal, due to the symmetry of the distribution
* The range of x is from -∞ to ∞
* The horizontal axis is asymptotic to the curve
* The total area beneath the curve is unity
* 68% of the values in a Normal distribution lie within ± 1 standard deviation of the mean
* 95% of the values lie within ±2 standard deviations of the mean
* 99.75% of the values lie within ±3 standard deviations of the mean

**X ~ N(μ, σ2)**

**The Standard Normal Distribution, Z ~ N(0, 12)**

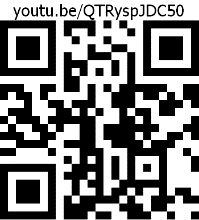
This is a Normal distribution centred at 0 with a variance and hence a standard deviation of 1. All Normal distributions with mean μ and variance σ2 can be adjusted to fit this curve as we will see later.



Even though the use of calculators make working with Normal distributions fairly straightforward, it is always a good idea to sketch a diagram to represent the probability you are trying to find.

Eg10 Find (a) P(Z < 1.52) (b) P(Z > 2.60) (c) P(Z < -0.75) (d) P(-1.18 < Z < 1.43)



Eg11 The random variable X ~ N(50, 42).

Find (a) P(X < 53), (b) P(X ≤ 45)

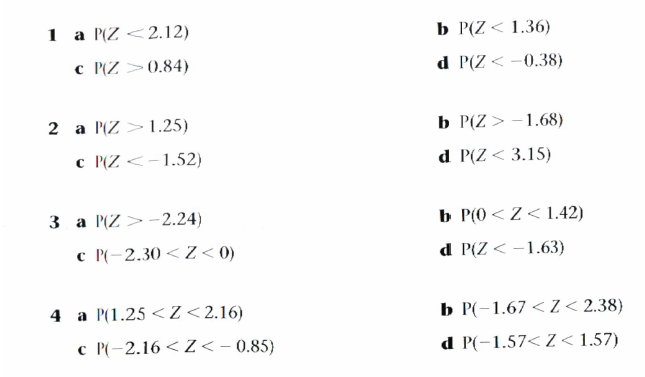
Eg12 When a butcher takes an order for a Christmas Turkey, he asks the customer what weight in kg the bird should be. He then sends his order to a turkey farmer who supplies birds of about the requested weight. For any particular weight of bird ordered, their error in kg may be taken to be normally distributed with mean 0 and standard deviation 0.75.

Mrs Jones orders a 10kg turkey from the butcher. Find the probability that the one she receives is

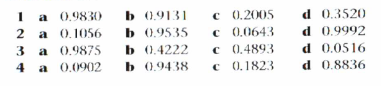
1. over 12kg
2. under 10kg
3. within 0.5kg of the weight she actually ordered.

**Exercise 2.3**

Find the following



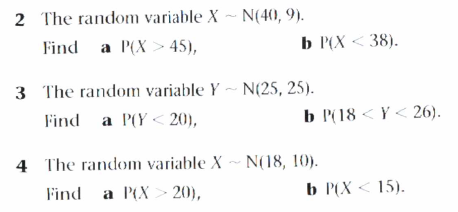
**Answers**



**Exercise 2.4**

1. The random variable X ~ N(30, 22)

Find (a) P(X < 33) (b) P(X > 26)



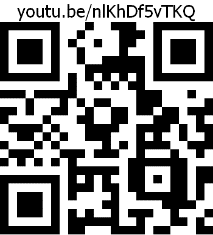


**Using a given probability to find the corresponding boundary parameter**

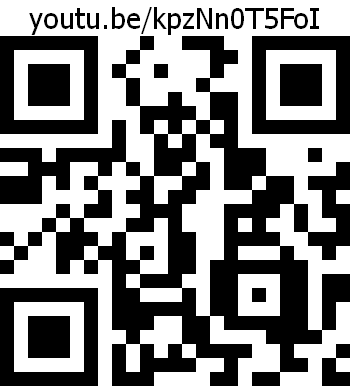
Once again, the calculator has made this a pretty straightforward process using the ‘Inverse Normal’ mode.

Eg13 Find the value of the constant *a* such that

1. P(Z < a) = 0.7611 (b) P(Z > a) = 0.01 (c) P(Z > a) = 0.0287 (d) P(Z < a) = 0.0170



Eg14 The random variable Y ~ N(20, 9). Find the value of b such that P(Y > b) = 0.0668

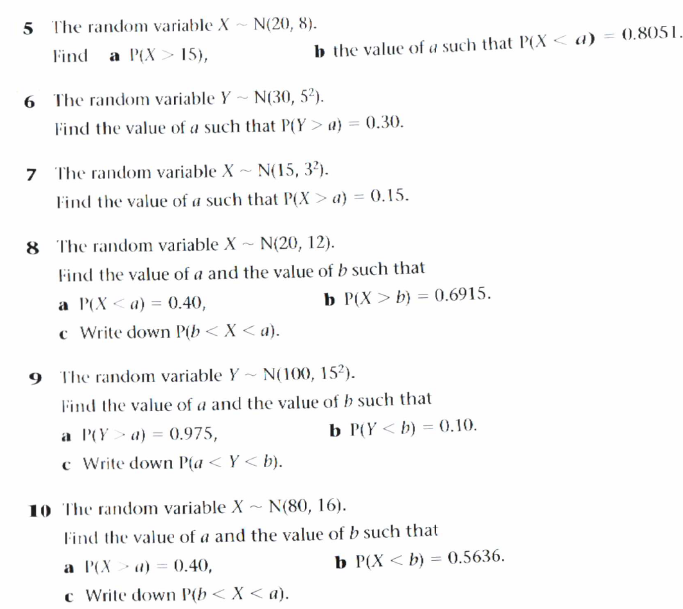


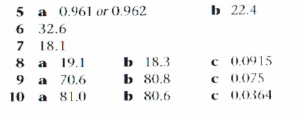
Eg15 In a particular experiment, the length of a metal bar is measured many times. The measured values are distributed approximately Normally with mean 1.340m and standard deviation 0.021m. Find the probabilities that any one measured value

1. exceeds 1.370m
2. lies between 1.310m and 1.370m
3. lies between 1.330m and 1.390m

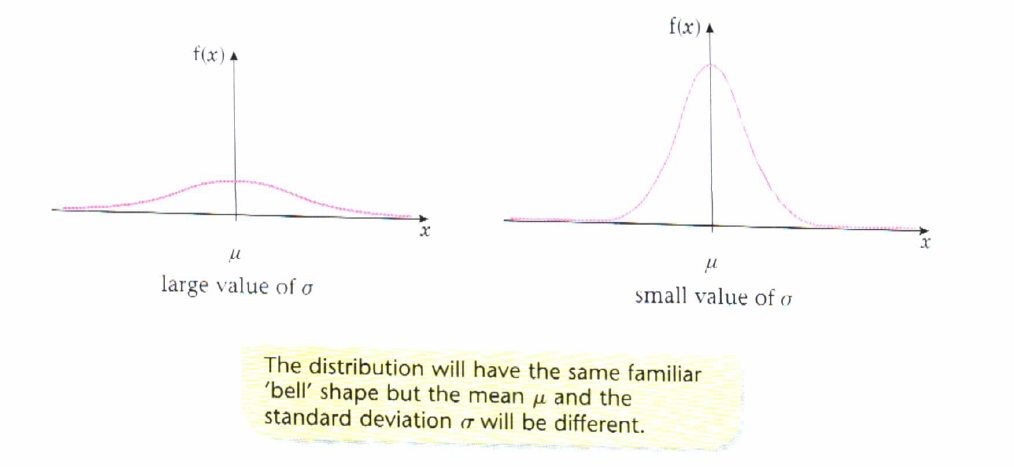
Find the length *l* for which the probability that any one measured value is less than *l* is 0.1.

**Exercise 2.5**





**Using the mappability of the Standard Normal Distribution, Z, to find the mean, standard deviation, or both for a Normal distribution, X**



X ~ N(μ, σ2) can be transformed into Z ~ N(0, 12) and vice-versa using the formula

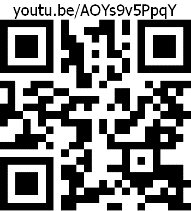
when tables rather than calculators were used to find probabilities of Normal distributions this had to be used to transform any given distribution into the Z distribution in order to read from the tables. This is no longer necessary. However, this process still needs to be used where situations call for you to determine an unknown mean or standard deviation value for a given distribution.

Eg16 The random variable X ~ N(μ, 32). Given that P(X > 20) = 0.20, find the value of μ.



Eg17 The random variable X ~ N(50, σ2). Given that P(X < 46) = 0.2119, find the value of σ.



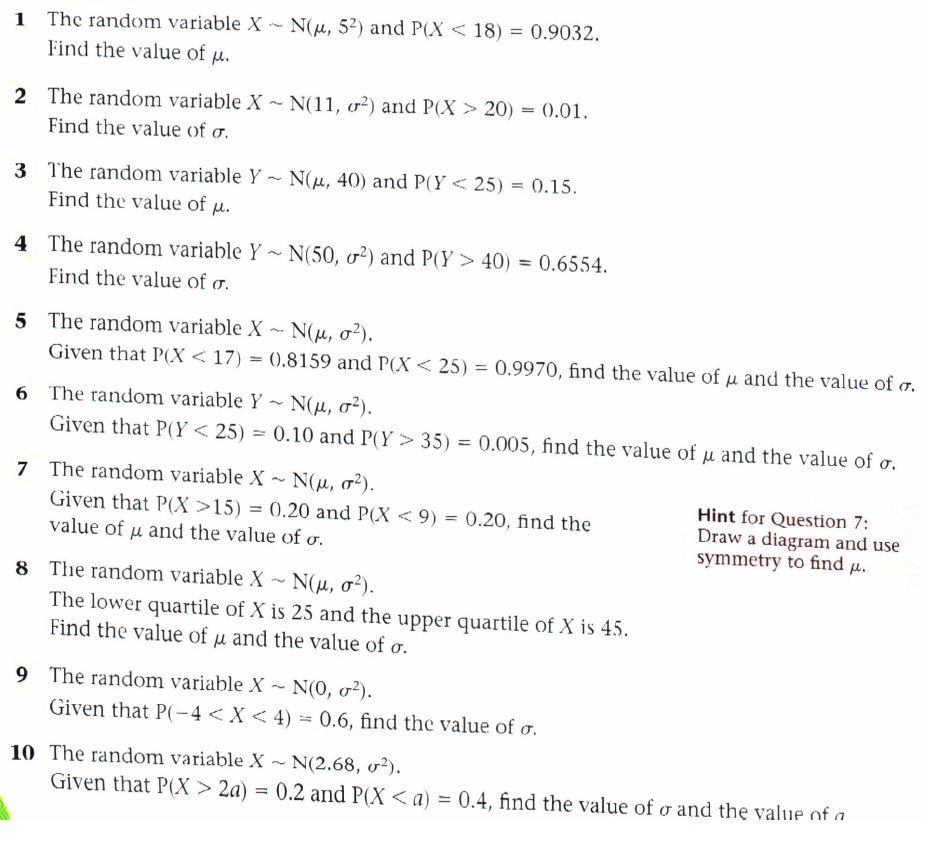
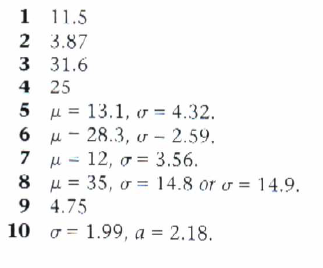
Eg18 The random variable X ~ N(μ, σ2). Given that P(X > 35) = 0.025 and P(X < 15) = 0.1469, find the mean and standard deviation of the distribution.

Eg19 A characteristic of the shape of a human skull is measured by a number *n*. People are classed into three groups: *A* (for which *n* ≤ 75), *B* (75 < *n* ≤ 80) and *C* (*n* > 80). In a certain population the percentages of people within these groups are 58, 38 and 4 respectively. Assuming that *n* is distributed Normally within this population, determine its mean and standard deviation.

 Three people are chosen at random from this population. Determine the probabilities that

1. each of the three has a value of *n* greater than 70;
2. at least one of the three has a value of *n* less than 70.

**Exercise 2.6**



**Problem Solving – Exercise 2.7**

