

Mathematics for Business and Information Systems

Dr Daniel Russell
Email: djrussell@kingston.ac.uk

1 Introduction

This handout provides support material for the module **BB4301: Business Information Analysis**. In the remainder of this section we provide motivation for studying mathematics on a business degree. This is followed by the following sections:

- 2 Sets
- 3 Functions and Relations
- 4 Propositional Logic
- 5 Predicate Logic
- 6 Descriptive Statistics
- 7 Probability
- 8 Probability Distributions

1.1 Motivation - Maths, Maths Everywhere

Since a very young age you have been learning and using mathematics. A lot of the songs that parents sing to young children involve numbers. For example,

One potato, two potato, three potato, four, five potato, six potato, seven potato, more!

and

One, two, three, four, five. Once I caught a fish alive. Six, seven, eight, nine, ten, then I let it go again. Why did you let it go? Because it bit my finger so. Which finger did it bite? This little finger on the right.

and even

There were ten in a bed and the little one said, "Roll over, roll over." So they all rolled over and one fell out. There were nine in the bed and the little one said, "Roll over, roll over." So they all rolled over and one fell out....There was one in the bed and the little one said, "Good night!"

Actually from the moment you are born you are described using mathematical terms. The child's weight, length, time of birth etc. are recorded, and in the case of weight, is always communicated to friends and relatives.

From maths-related nursery rhymes you move on at school to learn *inter alia*:

- the digits;
- simple arithmetic with the digits;
- numbers formed by more than one digit and more arithmetic;
- times tables;
- algebraic manipulation of expressions;
- writing and graphing functions;
- data handling and representation;
- geometry and trigonometry and so on.

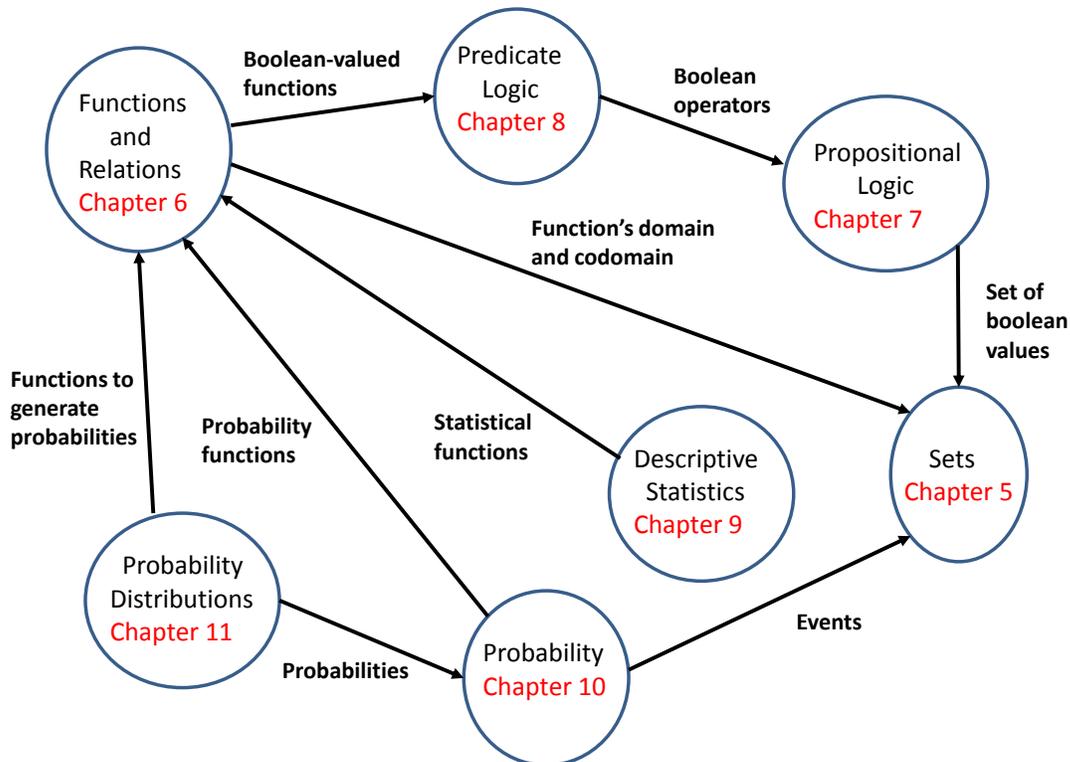
One thing they tend to forget to tell you when you learn mathematics is that it is not a collection of disparate topics but that each topic has links to one or more other topics. This is not surprising since mathematics was not discovered by a mathematics explorer, but has been developed over time using truths developed by previous mathematicians. For example, people often get anxious when asked to generate a **percentage**. But a percentage is just:

a name given to a **function** that is defined using the arithmetic operators \times and \div .

1. Write down an arithmetic expression to generate x as a percentage of y .
 2. Use the expression to evaluate the following:
 - (a) 10 as a percentage of 25;
 - (b) 60 as a percentage of 75;
 - (c) 30 as a percentage of 24.
-

On this module we will show as many of the dependencies between the various topics as possible, and further, will animate the material with examples from business and, more often, information systems. As a quick introduction to this animation, you have been using **sets, functions, propositional and predicate logic** a lot whenever you use a spreadsheet. **Relations** are essential to many popular database systems, and any business will need to summarize its financial status by ₂ using **descriptive statistics**.

Here is a pictorial representation of the dependencies between the various topics of this module. Each arrow goes from the dependent topic to the topic upon which it is dependent.



What is clear from this diagram is that all of the topics are dependent, directly or indirectly, on sets. For example:

- **descriptive statistics** is dependent on **functions** which are dependent on **sets**; and
- **probability distributions** are dependent on **probability** which is dependent on **sets**.

Before introducing the first topic of this module – **sets** – we present a more detailed illustration of the impact and influence of mathematics in business.

Spreadsheets are ubiquitous in business. They are used in most organisations to present, analyse and maintain data. A typical use is in processing sales data. That is, one can present the daily sales figures for various products, and produce summary **descriptive statistics** to present in a periodic report.

The sales figures will be presented using numeric values which can then be summarised using numeric descriptors such as the total, mean, standard deviation and so on. Now spreadsheet packages such as Excel categorise data into **types** where a type is simply a **set** of values with a common usage. All numeric values are members of the `Number` type.

Excel provides the user with a plethora of **functions** with which to manipulate data. We can use the `=SUM` **function** to generate the total sales figures for a week. This function only applies to₃ values of the `Number` type and returns a value of the same type.

It is often the case that a business will want to flag up figures that are significantly different than expected. This can be achieved by using the =IF **function** which tests its input and produces an output dependent on the result of the test. This testing will use boolean-valued **functions** from **predicate logic** and may use **propositional logic** operators.

In summary, a business application of a spreadsheet will use **sets** (the topic covered in the next chapter), **functions**, **descriptive statistics**, **propositional logic** and **predicate logic**!

1. Along with the `Number` type, what other types are provided by Excel?
 2. Name some Excel functions that require numbers as input.
 3. Name some Excel functions that require other values as input.
 4. What does the Excel =SUM function do with non-numeric values?
-

In the following section we introduce **sets** which are essential building blocks for many areas of mathematics.

2 Sets

In the previous section we described how the various topics that you have studied (or will study) in mathematics are inter-related and are used in various areas of business and information systems. The first topic of this module – **sets** – impact on all of the other topics. We begin with a formal definition of a set.

Definition 1 Set A set is an unordered collection of distinct objects.

Thus a set is an

- *unordered* – the order in which the elements of the set are written is not important;
- *collection of* – there are zero or more elements in a set;
- *distinct objects* – no element appears more than once.

Note that we refer to the objects in a set as the **elements** of the set.

A set is written between curly braces with the elements separated by commas. Thus, the set containing the elements 1, 2 and 3 is written

$$\{1, 2, 3\}$$

Of course it could also be written

$$\{2, 1, 3\}$$

since the order of the elements is not important.

2.0.1 Tuples

Sets are similar (but different) to **tuples** with which you are all familiar. You have for a long time described the coordinates of a point on a graph as a pair (2-tuple) with the first value representing the x-coordinate and the second representing the y-coordinate. For example,

$$(1, 2) \quad (-3, 8) \quad (0, 0)$$

Now a tuple is represented within parentheses and, in common with sets, the values are separated by a comma. Tuples however differ to sets in that there is an ordering. That is,

$$(1, 2) \neq (2, 1)$$

Also one is allowed repetition in a tuple. Hence on a 3-dimensional graph we have the following coordinates represented by triples (3-tuples).

$$(1, 1, 1) \quad (-2, 0, -2) \quad (5, 2, 2)$$

Finally, although sets typically include values of the same type, tuples often have mixed type values – with the caveat that each position must always have the same type. For example, a database table is a **set** of **records**. A record is simply a tuple of a particular size determined by the number of **fields** in the table. Each field has a particular type which may be numeric, textual, boolean and so on. Here is an elided database table represented as a list of tuples with the fields: a numerical *reference number*; *surname* of the text type; *initial* of the text type; *KU ID number* of the text type; and a numerical *office number*.

```
(1, Russell, D, ku02309, 4)
(2, Avery, B, ku12492, 340)
(3, Davis, K, ku35477, 381)
```

2.0.2 Back to Sets

We can make statements about the elements of a set. For example, if $A = \{5, 7, 9\}$ then

- $5 \in A$ is read *5 is an element of A*. This is clearly a true statement.
- $6 \notin A$ is read *6 is not an element of A*. This is also clearly a true statement.
- $8 \in A$ is however a false statement.

We can also make statements about the equality of sets. This requires a definition of set equality.

Definition 2 Set Equality Two sets are equal if they contain the same elements.

Thus, $\{3, 4, 5\} = \{5, 3, 4\}$ and $\{London, Paris, Rome, Madrid\} = \{Paris, Madrid, London, Rome\}$.

Note that since repetition is not allowed in sets, if two sets are equal then they must have the same number of elements. The number of elements in a set is called the **cardinality** of a set and is represented by two vertical lines enclosing the set. That is,

$$|\{1..10\}| = 10$$

Note that we use the $..$ notation as a shorthand representation of a sequence of values.

1. What is the minimum cardinality of a set?
 2. What is the maximum cardinality of a set?
 3. If two sets have equal cardinality does this imply that the sets are equal?
-

Before we describe how to combine sets we need to introduce two important sets – the **empty set** and the **universal set**.

- The empty set which is written $\{\}$ or \emptyset , is the set with no elements.
 - The universal set which is written U , is a set which contains all elements in a particular context. That is, it may be the (infinite) set of integers, the set of students on this module, or the set of countries in the Eurozone.
-

Use set notation to specify the following sets:

1. The set of positive even numbers less than 20.
 2. The set of vowels.
 3. The set of binary digits.
 4. The set of countries in the Eurozone.
 5. The set of states of America that begin with an M.
 6. The set of arithmetic operators.
 7. The set of planets in the solar system.
 8. The set of integers.
-

You should now be familiar with the characteristics and notation of sets. We now describe how to combine sets using set operators.

2.1 Operations on Sets

Learning about sets (and many other mathematical topics) is similar to learning **arithmetic**. First of all you learn about the objects – numbers in arithmetic and sets in this topic – and then you learn how to combine the elements. In arithmetic you learnt about the arithmetic operators:

- the **binary** (infix) operators $+$, $-$, \times and \div ; and
- the **unary** (prefix) operator $-$.

The binary operators, as their name suggests, take two numerical **operands** (inputs) and return a number. They are placed between their operands, hence the term **infix**. The unary operator takes only one numerical input and returns a number. It is placed in front of its input since it is a prefix operator.

Thus we can write the following arithmetic expressions

$$5 + 3 \quad (2.1)$$

$$7 - 12 \quad (2.2)$$

$$20 \div 8 \quad (2.3)$$

$$4 + 3 \times 7 \quad (2.4)$$

$$(3 - 6) \div 2 + 4 \times 5 \quad (2.5)$$

but the following expressions are invalid!

$$+53 \quad (2.6)$$

$$712 - \quad (2.7)$$

$$\div 208 \quad (2.8)$$

$$4 + \times 37 \quad (2.9)$$

$$\div (3 - 6)2 + 4 \times 5 \quad (2.10)$$

-
1. What other mathematical operators have you used?
 2. For each operator state whether it is **binary** or **unary**, and also, whether it is **infix** or **prefix**.
-

The set operators can also be categorized by their number of inputs. We have:

- the binary infix operators \cup (pronounced **union**), \cap (pronounced **intersection**) and \setminus or $-$ (pronounced **set difference**); and
- the unary operator \bar{A} or A^c (pronounced **complement** of A).

Here are the definitions of these operators.

Definition 3 Union The union of two sets A and B is the set that contains **all** of the elements in **either** set. It is written $A \cup B$.

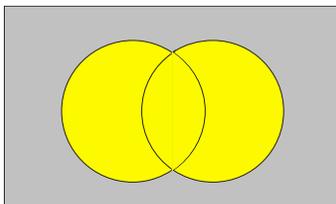
Definition 4 Intersection The intersection of two sets A and B is the set that contains **all** of the elements in **both** sets. It is written $A \cap B$.

Definition 5 Set Difference For two sets A and B , $A \setminus B$ (or $A - B$) is the set that contains **all** of the elements of A that are **not** in B .

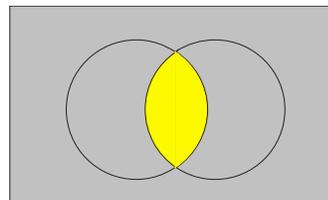
Definition 6 Set Complement The complement of set A is **all** the elements in U but not in A . It is written \bar{A} (or A^c) and is equivalent to $U \setminus A$ (or $U - A$).

Each of these operations may be represented pictorially using **Venn diagrams** which were introduced in the 19th century by John Venn. Each set is usually represented as a named ellipse (or circle) presented within a rectangle representing the universal set. The evaluation of an expression is typically shaded in the diagram. We present four example Venn diagrams as graphical representations of the four set operators. In our diagrams, where there are two sets, the set A is on the left, and the result of the operation is shaded in the lighter colour.

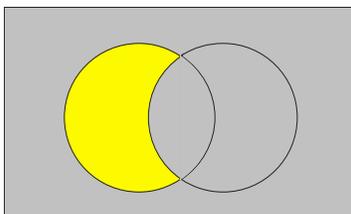
Union: $A \cup B$



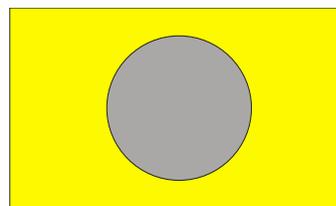
Intersection: $A \cap B$



Set Difference: $A \setminus B$



Complement: A^c



Thus we can write the following set expressions where A , B , C and D are sets.

$$A \cap B \quad (2.11)$$

$$\bar{B} \cup D \quad (2.12)$$

$$(A \setminus B) \cap C \quad (2.13)$$

$$(A \cup \bar{B} \cup C) \quad (2.14)$$

$$D \cap B \cup A \quad (2.15)$$

In the following section we describe how to evaluate – work out the value – of these expressions.

2.1.1 Evaluating Expressions

Of course, we can not only write arithmetic expressions but we can also **evaluate** them. We all learnt from a very young age how to add, subtract and even multiply and divide.

1. Evaluate expressions 2.1 to 2.5.
 2. What rules apply when evaluating arithmetic expressions?
-

2.1.2 BODMAS

We learnt from a young age that mathematical expressions are unambiguous. That is, they have one particular meaning, and when evaluated, result in one particular value. This requires some rules regarding how one evaluates an expression involving more than one operator. When evaluating arithmetic expressions we use **BODMAS** to determine the order of application of the operators in the expression. It is an acronym for

Brackets **O**f **D**ivide **M**ultiply **A**dd **S**ubtract

BODMAS states the **order of precedence** of the arithmetic operators (plus brackets and power).

- Brackets (or the expression within brackets) are always evaluated first.
- Any powers (this is what **O**f means) are evaluated next.
- Then divide and multiply are evaluated. They have the same precedence, so if they both appear they are evaluated from left to right.
- Finally add and subtract are evaluated. They also have the same precedence, so if they both appear they are evaluated from left to right.

This means that when confronted with the expression

$$(6 + 2)^2 - 10 \div 5 \times 2$$

it is evaluated as follows

$$\begin{aligned}(6 + 2)^2 - 10 \div 5 \times 2 &= 8^2 - 10 \div 5 \times 2 \\ &= 64 - 10 \div 5 \times 2 \\ &= 64 - 2 \times 2 \\ &= 64 - 4 \\ &= 60\end{aligned}$$

Evaluate the following expressions:

1. $6 \times (2 - 3)$
2. $5 + (4^2 - 1) \div (2 \times 5)$
3. $42 \div 2 - 2$
4. $42 \div (2 - 2)$

Now when evaluating set expressions we also have to abide by some rules. We will use the less friendly acronym **BCIUS** (pronounced *beesius*)!

BCIUS states the **order of precedence** of the set operators (plus brackets).

- Brackets (or the expression within brackets) are always evaluated first.
- Complements are evaluated next.
- Intersections are evaluated next.
- Finally unions and set differences are evaluated. They have the same precedence, so if they both appear they are evaluated from left to right.

This means that when confronted with the expression (assuming that the universal set is $\{1..10\}$)

$$(\{1, 2, 3\} \setminus \{3, 4, 5\})^c \cap (\{6, 5, 8\} \cup \{1, 6, 9\})$$

it is evaluated as follows

$$\begin{aligned}(\{1, 2, 3\} \setminus \{3, 4, 5\})^c \cap (\{6, 5, 8\} \cup \{1, 6, 9\}) &= \{1, 2\}^c \cap (\{6, 5, 8\} \cup \{1, 6, 9\}) \\ &= \{3..10\} \cap (\{6, 5, 8\} \cup \{1, 6, 9\}) \\ &= \{3..10\} \cap \{1, 6, 5, 8, 9\} \\ &= \{6, 5, 8, 9\}\end{aligned}$$

Let the set A be $\{0, 10\}$, B be $\{2, 4, 6, 8, 10\}$ and C be $\{-10..5\}$. Evaluate the following expressions:

1. $A \cap B$
 2. $(A \cup B) \setminus C$
 3. $(B \cap A)^c \cap C$
 4. $A^c \cap B^c \cap C^c$
 5. $(A \cup B \cup C)^c$
-

2.1.3 Laws of Set Algebra

The evaluation of set expressions (and arithmetic) is supported by a collection of laws. These are known as the **laws of set algebra**. These laws are presented in the following enumerated list where A , B and C are any sets. Where appropriate we will make reference to the equivalent arithmetic law.

1. $A \cup B = B \cup A$ union (in common with addition and multiplication) is **commutative**. This is because the union is simply all of the elements in either set and, therefore, the order of the operands is unimportant.
2. $A \cap B = B \cap A$ intersection (in common with addition and multiplication) is **commutative**. This is because the intersection is simply the elements that appear in both sets and, therefore, the order of the operands is unimportant.
3. $A \cup (B \cup C) = (A \cup B) \cup C$ union (in common with addition and multiplication) is **associative**. This is because the union of three sets is simply all of the elements in any of the sets. The order in which you apply the unions is unimportant.
4. $A \cap (B \cap C) = (A \cap B) \cap C$ intersection (in common with addition and multiplication) is **associative**. This is because the intersection of three sets is simply the elements that appear in all of the sets. The order in which you apply the intersections is unimportant.
5. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ union **distributes** over intersection. This is because $A \cup (B \cap C)$ represents all the elements of A plus any that are both in B and C , and $(A \cup B) \cap (A \cup C)$ represents elements that are both in $A \cup B$ and $A \cup C$. That is, the elements in A plus those that are in both B and C .
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ intersection **distributes** over union. Similar to the previous explanation. I'll leave this as a question.
7. $(A \cap B)^c = A^c \cup B^c$ **De Morgan's Law**. This is because $(A \cap B)^c$ represents the elements in the universal set that are not in both A and B , and $A^c \cup B^c$ represents elements that are in the universal set but are either not in A or not in B (that is, not in both). These are the same sets!
8. $(A \cup B)^c = A^c \cap B^c$ **De Morgan's Law**. Similar to the previous explanation. I'll leave this as a question.

9. $A \cup \emptyset = A$ \emptyset is the **identity** for union (zero is the identity for addition). This is because the empty set adds nothing to a union.
 10. $A \cap U = A$ the universal set is the **identity** for intersection (one is the identity for multiplication). This is because all elements of A are in U and they are the only elements they have in common.
 11. $A \cap A^C = \emptyset$ **complement law**. This is because, by definition, the two sets have nothing in common.
 12. $A \cup A^c = U$ **complement law**. This is because, by definition, every element in the universal set is either in a set or not.
-

1. Use a Venn diagram to show $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
 2. Use a Venn diagram to show $(A \cup B)^c = A^c \cap B^c$.
 3. Explain why subtraction is not commutative?
 4. Is subtraction associative?
 5. Either *addition distributes over multiplication* or *multiplication distributes over addition*. Which of these statements is true?
 6. Why is zero the identity for addition and one the identity for multiplication?
-

We can now use these laws and the definitions of the operators to evaluate some set expressions. Here are some examples.

In the first set of examples the universal set is the set of integers between -10 and 10 .

1. $\{1, 2, 3\} \cup \{4, 1, 6\} = \{1, 2, 3, 4, 6\}$
2. $\{4, 1, 6\} \cup \{1, 2, 3\} = \{1, 2, 3, 4, 6\}$ – same as above by commutative law
3. $(\{1..6\} \cap \{5..10\})^c = \{-10..4, 7..10\}$
4. $\{1..6\}^c \cup \{5..10\}^c = \{-10..4, 7..10\}$ – same as above by De Morgan's law.

In this second set of examples, let the universal set be all students in the Faculty, M be the set of students on this module, F be the set of female students in the Faculty and E be the set of 18 year old students in the Faculty.

1. $M \cap F$ is the set of female students on this module.
2. $M \cap (E \cup F)$ is the set of 18 year old or female students on this module.
3. $(M \cap E) \cup (M \cap F)$ is the same as above because intersection distributes over union.
4. $M \cup M^c$ is all the students in the Faculty by the complement law.
5. $E \cap (M \cap F^c)$ is all the 18 year old, male students on this module.

6. $(E \cap M) \cap F^c$ is the same as above by the associative law.

Let M be the set of Business and Law modules, L be the set of level 4 Business and Law modules and A be the set of Business and Law modules studied by an Accounting and Finance student. The universal set is the set of Kingston University modules. Evaluate the following set expressions.

1. $L \cap A$
 2. $M \cup A \cup L$
 3. $A \cap A^c$
 4. $A \setminus L$
 5. $L^c \cap A \cap M$
 6. $(A \cup L)^c$
 7. $A^c \cap L^c$
 8. $M \cup (L \cap L^c)$
-

An important relationship between sets is the **subset** relationship which we describe in the following section.

2.2 Subsets

We begin this section with the definition of a subset.

Definition 7 Subset A set A is a subset of set B , written $A \subseteq B$, if every element of A is in B .

Thus, the following sets are subsets of $\{1, 2, 3, 4\}$:

- $\{1, 3\}$
- $\{4, 1, 3\}$
- \emptyset

and $\{1, 2, 5\}$ is not a subset of $\{1, 2, 3, 4\}$, written $\{1, 2, 5\} \not\subseteq \{1, 2, 3, 4\}$, because $5 \notin \{1, 2, 3, 4\}$.

As another example the following are subsets of the set of Eurozone countries:

- $\{\text{France}\}$
- $\{\text{Italy, Netherlands, Germany, France}\}$

We create subsets of a set by removing **zero or more** elements from a set. This leads naturally to the followings subset laws:

- For any set A , $A \subseteq A$. That is, if zero elements are removed the set itself is created.
- For any set A , $\emptyset \subseteq A$. This is because if we remove all the elements from a set, the result is the empty set.
- For any sets A , B , and C , if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$. This is because, by definition, any element of A is an element of B and any element of B is an element of C . We say that the subset operator is **transitive**.
- For any sets A and B , if $A \subseteq B$ and $B \subseteq A$ then $A = B$. This is because, by definition, every element of A is an element of B and every element of B is an element of A .

Given the description of how subsets are created we know that if $A \subseteq B$ then $|A| \leq |B|$.

1. Which of the following statements is true?

- (a) $\{5, 6, 7\} \subseteq \{1..10\}$
- (b) If $|A| < |B|$ then $A \subseteq B$.
- (c) If $A \not\subseteq B$ and $B \not\subseteq C$ then $A \not\subseteq C$.
- (d) The empty set has no subsets.
- (e) $\emptyset = \{0\}$

2. List two other operators that are transitive.

Now we know how to create subsets, but how do we create all the subsets of a set and how many are there?

2.2.1 Power Set

We call the set of all subsets of a set A , the **power set** of the set A . This is written $P(A)$. Now each element of a set is either in a subset or not. That is, for each element there are two possibilities. For example, the set $\{1, 2\}$ has the following subsets:

- \emptyset when 1 **is not in** the subset and 2 **is not in** the subset;
- $\{1\}$ when 1 **is in** the subset and 2 **is not in** the subset;
- $\{2\}$ when 1 **is not in** the subset and 2 **is in** the subset;
- $\{1, 2\}$ when 1 **is in** the subset and 2 **is in** the subset.

So there are four subsets: 2 possibilities for the element 1 \times 2 possibilities for the element 2.

We can use Pascal's triangle for some insightful information about the subsets of a set.

2.2.2 Pascal's Triangle

Pascal's Triangle is named after the French mathematician Blaise Pascal. It is generated by putting the number 1 at the apex of the triangle and at the beginning and end of each row. Each row has one more element than the previous row. All other elements are generated by adding the two numbers to the upper left and upper right of the number. Here are the first five rows of Pascal's Triangle.

$$\begin{array}{rcccccc} n = 0: & & & & & & 1 \\ n = 1: & & & & 1 & & 1 \\ n = 2: & & & 1 & 2 & 1 & \\ n = 3: & 1 & & 3 & 3 & 1 & \\ n = 4: & 1 & 4 & 6 & 4 & 1 & \end{array}$$

Generate the next two rows of Pascal's Triangle.

Now interestingly each row provides information regarding the subsets of a set with the number of elements equal to the value of n . As you read across the row from left to right, the numbers represents the number of subsets with zero elements, 1 element, 2 elements and so on. The sum of the numbers in any row is the total number of subsets. For example:

- a set with zero elements (empty set) has 1 subset with zero elements;
- a set with one element has 1 subset with zero elements and 1 with one element. It thus has 2 subsets; and
- a set with two elements has 1 subset with zero elements, 2 with one element, and 1 with two elements. It thus has 4 subsets.

In summary a set with n elements has 2^n subsets.

1. How many subsets has a set with 5 elements.
 2. Explain why a set with n elements, has exactly the same number of subsets with 1 element as subsets with $n - 1$ elements.
-

We will use Pascal's Triangle again later in this module when studying **probability distributions**.

In the following section we introduce **functions** and **relations** which use sets in their definition, and are essential to several areas of mathematics, and many areas of business and information systems.

3 Functions and Relations

In everyday life we often encounter situations where an input is associated with one or more outputs. For example:

- we input a code into an ATM and are presented with an output screen. Thus the code is associated with an output screen;
- we input a telephone number into our mobile phone and are connected to another telephone. Thus the telephone number is associated with a telephone;
- we input a search word or sentence into Google and receive a list of search results. Thus a search word is associated with one or more search results;
- we input an ID number and password into the University system and are presented with a 'Welcome' screen. Thus an ID number/password combination is associated with a 'Welcome' screen;

For each of these examples, there are several (possibly infinite) input values, and several (possibly infinite) output values. The input values **must all be of the same type**:

- valid input codes for an ATM are all 4 digit values;
- valid telephone numbers are all a sequence of digits;
- Google search values are all text values; and
- there is a collection of valid ID numbers and matching valid passwords.

Each of these are **sets**: the set of ATM codes; the set of telephone numbers; the (infinite) set of text values; and the set of ID number/password values.

Equally the output values **must all be of the same type**:

- there is one output screen that everybody receives;
- all connections are made to telephones;
- Google searches always result in a list of search results; and
- there are a collection of user-specific 'Welcome' screens.

Each of these are **sets**: the (singleton) set of ATM welcome screens; the set of telephones; the (infinite) set of search results; and the set of user-specific screens.

Thus in each of these examples we have an **input set** of values, an **output set** of values, and a collection of associations between the input set and the output set.

This leads naturally to the following definition.

Definition 8 Relation A (binary) relation is a set of associations between elements of two sets.

Thus all of our examples are **relations**. On the ITB module last semester you were introduced to **relational** databases (such as Microsoft's Access), whose data model is built on (n-ary) relations. That is, a database table is a set of tuples whose size is determined by the number of fields. Each tuple represents an association between elements of n sets. Thus a database table is a set of associations between elements of n sets.

Now, there is a very important subset of relations called **functions**.

Definition 9 Function A function is a relation between a given set of elements called the domain and a set of elements called the codomain. The function associates each element in the domain with **exactly one element** in the codomain.

Here are some example functions. In each case we state the **domain** (input set), **codomain** (output set) and **the rule** that associates each element of the domain with exactly one element of the codomain. We have given each function a name that indicates what the function does – **this is good practice!**

Function Name	Domain	Codomain	Rule
double	integers	integers	double the input
capitalOf	countries	cities	return capital of a country
idNumber	students on this module	KU ID numbers	return id number of student
wordCount	books in the LRC	non-negative integers	number of words in book
celsToFahr	real numbers	real numbers	return temp in fahrenheit
always0	integers	integers	always return zero

When we discuss functions and their inputs we talk of a function **being applied to** its input. For example, the function `double` may be applied to any integer. We typically write the application of a function to its input with the function name appearing to the left of the input. The input is either separated from its input by parentheses or simply by using one or more spaces. For example,

`double(6)` `double 6` `capitalOf(Finland)` `capitalOf Finland`

Most mathematics texts and information systems software use the parenthesised approach, but there are some programming languages that use the space approach.

Each of the following are relations but not functions since each input value is not necessarily associated with exactly one value in the codomain.

Relation Name	Domain	Codomain	Rule
squareRoot	integers	real numbers	square root of number
borderingCountry	non-island countries	countries	country that borders the country
moduleStudied	students on this module	KU modules	return modules studied by student

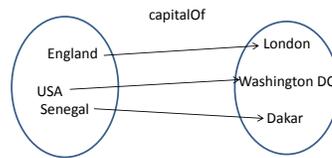
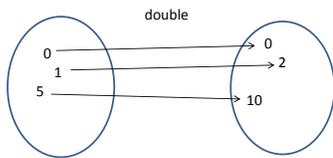
1. Which of the following is a relation and which is a function?

(a) `areaOfCircle` that associates a radius with an area.

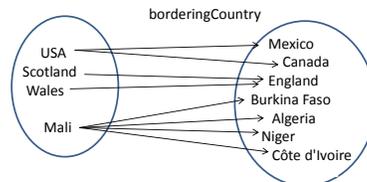
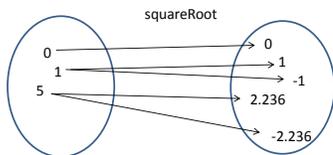
- (b) `sittingNextTo` that associates a student in this class with a student they are sitting next to.
- (c) `age` that associates a student on this module with their age.
- (d) `dependentMathsTopic` that associates a maths topic with a topic it is dependent on.
- (e) `ipAddress` that associates an internet protocol address with a device.
- (f) `sum` that associates a list of numbers with its sum
- (g) `powerSet` that associates a set with its power set

2. For each function, state its domain and codomain.

In common with sets, and most areas of mathematics, it is useful to present functions using a pictorial representation. At school you will have seen many numerical functions (such as `double`) presented as a straight line on a graph. We will be using a different notation, where the domain and codomain are presented as ellipses, and the associations as a collection of arrows from the domain to the codomain. We present the functions `double` and `capitalOf` using this notation.



We can of course represent relations using this same notation. For example, here are the relations `squareRoot` and `borderingCountry`. In each case we have only presented a small subset of the domain and codomain in the picture.



An important subset of the codomain of a function is the **range** of a function. It is the set of elements of the codomain that are associated with elements in the domain. For example:

- the range of `double` is the set of even integers;
- the range of `capitalOf` is the set of capital cities;
- the range of `wordCount` is the set of integers that represent the number of words in each book of the library; and
- the range of `celsToFahr` is all the real numbers (although in reality temperatures have a finite range).

As has previously been stated functions are ubiquitous in mathematics, business and information systems. This is because, by definition, for any input, you know that there is only one output. So all of the functions in a spreadsheet like Excel return exactly one output for any input. We now list some Excel functions with their domain, codomain and rule.

Excel Function	Domain	Codomain	Rule
=SUM	lists of numbers	numbers	sum the numbers
=CONCATENATE	lists of text	text	concatenate the text
=AND	lists of boolean values	boolean values	returns TRUE if all TRUE

What is the range of each of the following functions?

1. The function `square` that squares its input. The domain and codomain is the integers.
 2. The function `always0` that always returns zero. The domain and codomain is the real numbers.
 3. The function `isEven` that returns true if the input is even and false otherwise. The domain is the integers and the codomain the boolean values.
 4. The function `idNumber` that returns the ID number of a student. The domain is the set of students on this module and codomain is the set of KU ID numbers.
-

An alternative way to represent a relation or a function is as a **set of ordered pairs**.

3.1 Cartesian Product and Ordered Pairs

The Cartesian product of two sets A and B written $A \times B$ is the set of all possible **ordered pairs** where the first element is from A and the second element is from B . Cartesian products are named after René Descartes, a French mathematician and philosopher. Here is an example,

$$\{1, 2\} \times \{0, 4, 8\} = \{(1, 0), (1, 4), (1, 8), (2, 0), (2, 4), (2, 8)\}$$

Note that for many years you have (probably unknowingly) been using Cartesian products. Whenever you have created a graph on graph paper you have been using the **Cartesian plane**, where each point on the graph is represented by a **Cartesian coordinate** such as $(1, 0)$ and $(-3, 6)$. The Cartesian plane is the set of ordered pairs generated by the Cartesian product of the real numbers \mathbb{R} with themselves. That is, $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Now any graph is simply a subset of \mathbb{R}^2 and, in general, any relation between A and B is a subset of $A \times B$. For example, the relation `lessThan` may be represented as the following set of ordered pairs.

$$\text{lessThan} = \{(1, 4), (1, 8), (2, 4), (2, 8)\}$$

Further, any function between A and B is also a subset of $A \times B$. For example, `multByFour` may be represented as the following set of ordered pairs.

$$\text{multByFour} = \{(1, 4), (2, 8)\}$$

For an ordered pair that represents a function **each element from the domain set appears only once**.

-
1. Represent each of the following relations and functions as a set of ordered pairs.
 - (a) `capitalOf`
 - (b) `cityIn` that associates a country with a city in the country
 - (c) `areaOfCircle`
 - (d) `powerSet` that associates a set with its power set
 - (e) `isEven` which associates a number with `True` if its is even and `False` otherwise
 2. For any two sets A and B , what is the relationship between $|A|$, $|B|$ and $|A \times B|$.
-

In the following section we introduce an operator that enables the construction of new functions from existing functions.

3.2 Function Composition

In common with numbers and sets, functions may be combined. There is a single binary infix operator called the **function composition operator** which is denoted \circ .

If f and g are functions, then the expression $f \circ g$ is read **f composed with g**.

This means **apply g and then apply f to the result of g**. This appears counter-intuitive in that g is applied before f , that is, they are not applied from left to right. The reason, they are applied in this order, is that when you write the expression with an input, say x then it is g that appears closest to the input. That is, $(f \circ g)(x)$.

Thus we can create new functions from existing functions simply by gluing them together using the function composition operator. For example:

- the function `quadruple` can be created by composing `double` with itself;
- the function `areaOfCircle` can be created by composing the function `square` with the function `multByPi`; and
- the function `percentage` can be created by composing `divide` with `multBy100`.

Now, to compose f with g as above, the codomain of g must be a subset of the domain of f . That is, any output of g must be a valid input of f .

Here are some example applications of function composition.

- $(\text{double} \circ \text{addOne})(5) = 12$ where `addOne` is a function that adds one to an integer.
- $(\text{addOne} \circ \text{double})(5) = 11$ this illustrates that function composition is **not commutative**.
- $(\text{convertToNumber} \circ \text{monthOfBirth})(\text{ku02309}) = 5$ where `monthOfBirth` takes a KU ID number and returns the person's month of birth, and `convertToNumber` converts a month to a number indicating the position of the month in the year.

- $(\text{double} \circ \text{numberOfVowels})(\text{Maths}) = 2$ where `numberOfVowels` returns the number of vowels in a word.
 - $(\text{numberOfVowels} \circ \text{double})(6)$ this does not work since the codomain of `double` is not a subset of the domain of `numberOfVowels`.
-

1. Give sensible names to the functions created by the following compositions.

- `addOne` \circ `double`
- `double` \circ `addOne`
- `add32` \circ `multBy9` \circ `divideBy5`
- `motherOf` \circ `motherOf`
- `firstSisterOf` \circ `motherOf`
- `motherOf` \circ `firstSisterOf`
- `cardinality` \circ `powerSet`

2. Evaluate the following expressions.

- $(\text{addOne} \circ \text{double})(10)$
- $(\text{double} \circ \text{addOne})(10)$
- $(\text{add32} \circ \text{multBy9} \circ \text{divideBy5})(10)$
- $(\text{motherOf} \circ \text{motherOf})(\text{you})$
- $(\text{firstSisterOf} \circ \text{motherOf})(\text{you})$
- $(\text{motherOf} \circ \text{firstSisterOf})(\text{you})$
- $(\text{cardinality} \circ \text{powerSet})(\{1..6\})$

3. Is function composition associative?

4. Why are function composition expressions placed in parentheses when applied to an input?

In the following section we describe some categorisations of functions.

3.3 1-1, Onto and Inverse Functions

We begin this section with the definitions of two characteristics of a function.

Definition 10 1-1 Function A function is 1-1 (also known as **injective**) if each element of the domain is associated with a different element of the codomain. That is, if `fun` is a 1-1 function, then for any two inputs x and y , $\text{fun}(x) \neq \text{fun}(y)$. Note that if a function is not 1-1 then it is **many-1**.

Definition 11 Onto Function A function is onto (also known as **surjective**) if each element of the codomain is associated with at least one element of the domain. That is, if `fun` is an onto function, then for any element q in the codomain, there exists an element p in the domain, such that $\text{fun}(p) = q$.

1-1 functions are very important in everyday life, information systems and business, since they guarantee that no two inputs may have the same output. Here are some examples of functions that **must** be 1-1:

- the function that assigns an ID number to each Kingston University student. Each student must be uniquely identified by their ID number;
- the function that takes a primary key value and returns a record in a database table. By definition, a primary key value must identify a single record in a table;
- the function that associates a Dewey decimal code with a library book.

Onto functions are also important in that they guarantee that every possible output value can be achieved. Here are some examples of functions that must be **onto**:

- the function that associates a ID number with a student on the module;
- the function that associates a National Insurance number with a person over the age of 16 born and resident in the UK; and
- a function that associates a capital city with a country in the world.

It is important to note that determining whether a function is 1-1 and/or onto, is not solely dependent on the rule, but also relies on the domain and codomain of the function. For example, the function *double* from integers to integers is **not onto** since odd numbers are never returned. However, if we change the codomain to the even integers then the function is onto.

-
1. Which of the following functions is 1-1, onto, both 1-1 and onto or neither 1-1 nor onto.
 - (a) the function *double* with domain and codomain the integers.
 - (b) the function *celsToFahr* with domain and codomain the real numbers.
 - (c) the function *isEven* with domain the integers and codomain the boolean values true and false.
 - (d) the function *cardinality* with domain the set of sets and codomain the non-negative integers.
 - (e) the function *age* with the domain the students on this module and codomain the set $\{18..35\}$.
 - (f) the function *isPositive* from the set $\{1, -1\}$ and the codomain the boolean values.
 2. If two 1-1 functions are composed, is the resulting function 1-1?
 3. If two onto functions are composed, is the resulting function onto?
-

3.3.1 Cardinality, 1-1 and Onto

Here are some rules relating the cardinality of the domain and codomain of a function, and the classification of a function. In all of these rules the domain is called D and the codomain C .

- If $|D| > |C|$ then the function cannot be 1-1.
- If $|D| < |C|$ then the function cannot be onto.
- If a function is 1-1 and $|C| = |D|$ then the function must be onto.
- If a function is onto and $|C| = |D|$ then the function must be 1-1.

Now functions that are both 1-1 and onto have an **inverse function**. This is exactly what it says on the tin: it does the reverse of the original function. For example: the inverse of `double` is `halve` and the inverse of `capitalOf` is `countryIn`. Each of these functions can be pictorially represented simply by swapping the domain and codomain of the original function, and turning each arrow around.

For a function to have an inverse it must be:

- 1-1, since if it were not, when the function was applied in reverse there would not be a single output associated with each input; and
- onto, since if it were not, when the function was applied in reverse there would be elements in the domain with no association.

Definition 12 Inverse Function The inverse function of a 1-1 and onto function f , denoted f^{-1} , is defined as the function such that for any element a in the domain of f , if $f(a) = b$ then $f^{-1}(b) = a$. Thus if we compose f with its inverse we get an **identity** function.

Definition 13 Identity Function A function f whose domain is the same as the codomain, and for any element a in the domain $f(a) = a$.

The following are examples of identity functions created by the composition of a function with its inverse.

- $(\text{halve} \circ \text{double})(7) = 7$
- $(\text{countryIn} \circ \text{capitalOf})(\text{Brazil}) = \text{Brazil}$
- $(\text{capitalOf} \circ \text{countryIn})(\text{Rome}) = \text{Rome}$

-
1. Give sensible names to the inverse function of each of the following functions.
 - (a) `addOne`
 - (b) `celsToFahr`
 - (c) `reverse` a function which reverses a piece of text
 - (d) `swapCoordinates` a function which reflects a point in the Cartesian plane about the line $x = y$
 - (e) `tomorrow`
 2. Explain why the function `cardinality` that returns the number of elements in a set does not have an inverse.
 3. What is the inverse of the inverse of a function?
-

The next two sections cover **logic** an essential tool in the development of mathematics.

4 Propositional Logic

In last semester's Information Technology for Business (ITB) module, you studied databases and a query language for accessing data from a database. The query language SQL enables the user to select data from named fields in named tables that satisfy some explicit conditions required of one or more fields. These conditions, typically written in a `WHERE` clause, are a statement whose value is true for the required fields. Here are some examples from the ITB module.

```
SELECT *
FROM emp
WHERE mgr = 734;
```

This query returns all of the fields of the records in table `emp` for which the manager code is 734.

```
SELECT *
FROM emp
WHERE job = "Admin" and sal > 22000;
```

This query returns all of the fields of the records in table `emp` for which the job is administrative and the salary is greater than £22000.

In general, there types of SQL queries can be described as

Get one or more fields from those records in a table that satisfy some explicit condition(s).

Now the **condition(s)** are stated in the `WHERE` clause and its value must be either True or False. That is, the value of `mgr = 734` is either true or false dependent on the value of `mgr`. If it is true for a particular record then all the fields for that record is returned by the query. If it is false, then nothing is returned.

The value of `job = "Admin" and sal > 22000` is true if **both** `job = "Admin"` is true and `sal > 22000` is true.

Note that each field has a particular **type**. That is, the values in every record for a particular field must come from the same set of values. Thus sets have an impact on the structure of a database record.

Note also that:

- True and false are the **logical** or **truth** values. They are represented by the Boolean type.
- `and` is one of the logical operators.

The area of mathematics that studies truth values and their combination is **logic**. This includes various sub-topics including the two we will study on this module:

- **propositional logic** which is the study of truth-valued entities (**propositions**) and the effects of combining them using various operators; and
- **predicate logic** which is the study of truth-valued functions (**predicates**) and the effects of combining them using various operators.

We will first study propositional logic which requires a formal definition of a proposition.

4.1 Propositions

Definition 14 Proposition A (declarative) sentence that expresses something that is either true or false.

In everyday conversation we often communicate using propositions. For example:

- My name is Dan Russell.
- My ID number is KU02309.
- A set is an unordered collection of distinct objects.
- I am six feet tall.
- There are nine planets in the Solar System.

The common feature for all of these statements is not that they are all true, but that they all express something that is either true or false. None of them rely on opinion – they are known to be true or false.

They therefore differ from other sentences which are typical (and not so typical) in conversation, that include:

- Turn off your mobile phone.
- What is 6^3 ?
- `sal > 22000`

The first sentence is not a declarative sentence – it is a command. The second sentence is also not a declarative sentence – it is a question. The third sentence which was used in the SQL examples needs a value for `sales` to determine if it is true or false. We will return to this example in section 5.

Which of the following sentences is a proposition? Give reasons for your answer.

1. Rome is the capital of Spain.
 2. Strictly Come Dancing is better than X Factor.
 3. The empty set is a subset of any set.
 4. All relations are functions.
 5. $x^2 + y^2 = z^2$
 6. The binary number 01101111 is equal to the decimal number 111.
-

In common with sets and numbers, once we have learnt about the objects of the topic we need to learn how to combine these objects. In this case, how to create new, more complex propositions, from simple (atomic) propositions.

4.2 Combining Propositions

Propositions may be combined using **logical operators**.

In common with arithmetic and set operators, the logical operators can also be categorized by their number of inputs. We have:

- the binary infix operators: \vee (pronounced (inclusive) **or**), \wedge (pronounced **and**), \Rightarrow (pronounced **implies**) and \Leftrightarrow (pronounced **if and only if**); and
- the unary operator \neg (pronounced **not**).

Here are the definitions of these operators.

Definition 15 Or For any two propositions P and Q , $P \vee Q$ is true if **at least one** of them is true. Thus, $P \vee Q$ is true if:

- only P is true;
- only Q is true; or if

- both are true.

The only time it is false is if both are false. This is why it is called **inclusive or** since it includes the case when both are true. There is a less commonly used operator **exclusive or** which is written \oplus which is true when exactly one is true.

Definition 16 And For any two propositions P and Q , $P \wedge Q$ is true if **both** are true.

Definition 17 Implication For any two propositions P and Q , $P \Rightarrow Q$ is true if both are true, or if P is false. That is, the statement is true if the truth of P (called the **antecedent**) implies the truth of Q (the **consequent**). If P is false then the expression is true whatever the value of Q .

Definition 18 Bi-Implication (If and Only If) For any two propositions P and Q , $P \Leftrightarrow Q$ is true if P and Q have the same truth values. That is, they are both true or both false.

Definition 19 Negation (Not) For any proposition P , the value of $\neg P$ is the opposite of the value of P . That is, if P is true then $\neg P$ is false and if P is false then $\neg P$ is true..

Thus we can write the following propositional logic expressions where A , B , C and D represent propositions. They are formally known as **propositional variables** and can be replaced by any proposition (in the way that A and B can be replaced by any set in $A \cap B$, and x , y and z can be replaced by any number in $x + y + z$).

$$A \vee B \quad (4.1)$$

$$\neg B \wedge D \quad (4.2)$$

$$A \Rightarrow B \quad (4.3)$$

$$\neg(A \wedge B \wedge C) \quad (4.4)$$

$$(D \Leftrightarrow B) \vee (A \Leftrightarrow C) \quad (4.5)$$

The following statements are all true.

- Rome is the capital of Italy \wedge 7 is a prime
- Rome is the capital of Spain \vee 7 is a prime
- Rome is the capital of Italy \Rightarrow 7 is a prime
- Rome is the capital of Spain \Rightarrow 8 is a prime
- Rome is the capital of Spain \Leftrightarrow 8 is a prime

The following statements are all false.

- Rome is the capital of Spain \wedge 7 is a prime
- Rome is the capital of Spain \vee 8 is a prime
- Rome is the capital of Italy \Rightarrow 8 is a prime

- Rome is not the capital of Spain \Rightarrow 8 is a prime
- Rome is the capital of Spain \Leftrightarrow 7 is a prime

1. Evaluate the following logical expressions (state whether they are True or False).
 - (a) The empty set is a subset of every set \wedge a function is a relation
 - (b) The set $\{1,2,3\}$ has 7 subsets \Rightarrow Ouagadougou is the capital of Burkina Faso
 - (c) $\neg(\neg$ The Excel expression `=SUM(1,2,3,4,False,5)` evaluates to 15)
 - (d) The empty set has no subsets \Leftrightarrow function composition is commutative
2. Which of the logical operators are **commutative**?
3. Which of the logical operators are **associative**?

4.3 Negating an Expression

The application of the negation operator to a propositional logic expression results in

- each \wedge becoming an \vee ; and
- each \vee becoming an \wedge .

That is,

- $\neg(P \wedge Q)$ is true when **not both** are true, which is equivalent to **at least one is not** true. That is, $\neg P \vee \neg Q$; and
- $\neg(P \vee Q)$ is true when **not either (neither)** are true, which is equivalent to **both are not** true. That is, $(\neg P \wedge \neg Q)$.

We have seen this relationship previously in section 2.1.3 where we presented the Laws of Set Algebra, and specifically De Morgan's Laws. There, we described the result of the application of the **complement** operator to a **union** or **intersection**. Here we are presenting the effect of the **negation** operator when applied to an **or** or an **and**. It is not too difficult to understand that these are similar in their nature.

Here are some example applications of the negation operator. We will let R represent the proposition Rome is the capital of Italy and P represent the proposition 3 is a prime number. Then:

- $R \wedge P$ represents Rome is the capital of Italy and 3 is a prime number which is a true statement. When negated this becomes $\neg(R \wedge P)$ which is equivalent to $\neg R \vee \neg P$ and is translated to Rome is not the capital of Italy or 3 is not a prime number. This is false.

- $R \vee P$ represents Rome is the capital of Italy or 3 is a prime number which is a true statement. When negated this becomes $\neg(R \vee P)$ which is equivalent to $\neg R \wedge \neg P$ and is translated to Rome is not the capital of Italy and 3 is not a prime number. This is false.

In the following section we describe how to evaluate – work out the value – of propositional logic expressions.

4.4 Truth Tables and Order of Precedence

The value of a propositional logic expression depends on the values of its propositional variables. For example, the value of $P \wedge Q$ depends on the value of P and the value of Q . If P is replaced by 3 is a prime and Q is replaced by London is the capital of England then $P \wedge Q$ is true. But if Q is replaced by The empty set has 2 subsets then $P \wedge Q$ is false.

Thus the expression does not have a single value – it has a value for each combination of values for the propositional variables. Note that this is the same for:

- $m + n$ whose value depends on the numerical values of m and n ; and
- $S \cap T$ whose value depends on the set values of S and T .

Now each propositional variable has two possible values – True and False – and so the number of combinations of values is 2^n where n is the number of propositional variables.

Thus:

- an expression with two propositional variables has four combinations of values; and
- an expression with three propositional variables has eight combinations of values.

Note that for a simple arithmetic expression like $x + y$, if the domain is the integers, then there are **infinitely** many combinations of values for x and y and not just four!

A simple and clear way to present the possible values of a propositional expression is with a **truth table**. This is, as its name suggests, a table that presents the value of an expression for each possible combination of values of its proposition variables. We present the truth table for the expression $P \wedge Q$.

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

Note that T represents true and F represents false. The first two columns present the possible combinations of truth values for the **propositional variables** P and Q . The final column presents

the value of the expression for each combination. In this example, the expression is true only if the proposition variables are true.

A second example truth table presents the values of the expression $(\neg P \vee \neg Q \vee \neg R) \Leftrightarrow \neg(P \wedge Q \wedge R)$.

P	Q	R	$\neg P$	$\neg Q$	$\neg R$	$\neg P \vee \neg Q$	$\neg P \vee \neg Q \vee \neg R$	$P \wedge Q$	$P \wedge Q \wedge R$	$\neg(P \wedge Q \wedge R)$	$(\neg P \vee \neg Q \vee \neg R) \Leftrightarrow \neg(P \wedge Q \wedge R)$
T	T	T	F	F	F	F	F	T	T	F	T
T	T	F	F	F	T	F	T	T	F	T	T
T	F	T	F	T	F	T	T	F	F	T	T
T	F	F	F	T	T	T	T	F	F	T	T
F	T	T	T	F	F	T	T	F	F	T	T
F	T	F	T	F	T	T	T	F	F	T	T
F	F	T	T	T	F	T	T	F	F	T	T
F	F	F	T	T	T	T	T	F	F	T	T

In this example there are three proposition variables P , Q and R . There are therefore eight possible combinations of values for these propositions, and hence eight rows in the table. The first three columns of the table presents these combinations. The value of the expression is then determined by evaluating its sub-expressions using the appropriate rules of precedence.

4.4.1 Evaluation Order

In arithmetic we apply **BODMAS**, in set algebra we apply **BCIUS** and in Boolean algebra (the algebra used with truth-valued entities) we apply **BNAOIF** (not so easy to pronounce)! This acronym stands for:

- **B**rackets (or the expression within brackets) are always evaluated first.
- **N**ot is evaluated next. Then
- **A**nd. Then
- **O**r. Then
- **I**mplies. And finally
- **I**F and only if.

Thus $(\neg P \vee \neg Q \vee \neg R) \Leftrightarrow \neg(P \wedge Q \wedge R)$ is evaluated as in the illustration below. We have used the third row of the truth table in this example

$$\begin{aligned}
(\neg True \vee \neg True \vee \neg False) &\Leftrightarrow \neg(True \wedge True \wedge False) \\
&= (False \vee False \vee True) \Leftrightarrow \neg(True \wedge True \wedge False) \\
&= (False \vee True) \Leftrightarrow \neg(True \wedge True \wedge False) \\
&= True \Leftrightarrow \neg(True \wedge True \wedge False) \\
&= True \Leftrightarrow \neg(True \wedge False) \\
&= True \Leftrightarrow \neg False \\
&= True \Leftrightarrow True \\
&= True
\end{aligned}$$

The truth table mimics this evaluation in the columns between the first three and the final column.

Note that the final column contains only true values. This indicates that this expression is a tautology.

1. Write truth tables for each of the following expressions.

- (a) $P \neg P$
- (b) $\neg(P \vee Q)$
- (c) $P \Rightarrow (Q \vee P)$
- (d) $(P \Leftrightarrow Q) \Leftrightarrow R$

2. If an expression is satisfiable, how is this indicated in a truth table?

3. If we have a truth table for an expression, how do we generate the truth table for the negation of the expression?

4.5 Tautology and Contradiction

There are two categories of expressions that deserve some special focus: **tautologies** and **contradictions**.

Definition 20 Tautology A tautology is a propositional logic expression that is always true. That is, for any combination of truth values of the proposition variables the expression is true.

Definition 21 Contradiction A contradiction is a propositional logic expression that is always false. That is, for any combination of truth values of the proposition variables the expression is false.

Note that a propositional logic expression is **satisfiable** if there is a combination of values of its propositional variables for which the expression is true. By definition, a tautology is satisfiable and a contradiction is not satisfiable.

The following expressions are example tautologies.

- $P \vee \neg P$
- $Q \Rightarrow Q$
- $(Q \Rightarrow R) \Leftrightarrow (\neg Q \vee R)$
- $\neg(P \wedge Q) \Leftrightarrow (\neg P \vee \neg Q)$

and the following expressions are example contradictions.

- $P \wedge \neg P$
- $Q \Leftrightarrow \neg Q$

1. Let P and Q be propositional variables. Classify each of the following expressions as a **tautology**, **contradiction** or **satisfiable**. Give reasons for your answers.

- (a) $P \wedge Q$
- (b) $P \Leftrightarrow P$
- (c) $(P \wedge \neg P) \Rightarrow Q$
- (d) $(Q \vee P) \wedge \neg(Q \wedge P)$

2. If a proposition P is not satisfiable, what can you say about the propositional expression $P \wedge Q$ for any proposition Q ?

3. If a proposition P is not satisfiable, what can you say about the propositional expression $P \Rightarrow Q$ for any proposition Q ?

In the next section we discuss the second logic topic covered on this module – predicate logic.

5 Predicate Logic

In the previous section we used the following example as a motivation for studying logic.

```
SELECT *
FROM emp
WHERE mgr = 734;
```

We also explained that $\text{mgr} = 734$ is not a proposition since its truth value depends on the value of mgr . We call $\text{mgr} = 734$ a **predicate**.

Definition 22 Predicate A predicate is a **boolean-valued** function. Sometimes called a **propositional function**, it associates a truth value with an input value.

Thus the predicate $\text{mgr} = 734$ associates a value for mgr with true or false. That is, if the manager code is 734, then the predicate returns true. If the manager code is not 734 then it returns false. We give predicates names in the same way we give functions names. For example, $M(x)$ could represent $x = 734$ where x takes its values from the mgr values in the table emp .

Note that the definition of the predicate (as with any function) is dependent on the definition of its domain. For an SQL query the domain will be the values of the field(s) referred to in the predicate used in the query.

5.1 Quantifiers

With predicates we are able to introduce the idea of quantity into our logical expressions. That is, we can talk about **all values satisfying some condition** or **at least one value satisfying a condition**. These statements are expressed by using predicates associated with one or more **quantifiers**.

Definition 23 Universal Quantifier The universal quantifier \forall is pronounced **all, every** or **for all**.

Definition 24 Existential Quantifier The existential quantifier \exists is pronounced **there is, there exists** or **there is at least one**.

We illustrate the use of these quantifiers with some examples. Let $E(x)$ represent x is even, $P(y)$ represent y is positive and $G(s, t)$ represent s is greater than t . The domain for these predicates is the integers.

Here are some example expressions and their translation into English. Note that in each example a quantifier is associated with a variable. Formally this converts a **free variable** into a **bound variable** – in this case bound to the set of values specified by the domain. Any valid predicate logic expression must include only bound variables.

- $\forall m E(m)$ translates to **all integers are even**.
- $\exists m P(m)$ translates to **there is a positive integer**.
- $\forall x \neg E(x)$ translates to **all integers are not even** or **all integers are odd**.
- $\exists x \exists y G(x, y)$ translates to **there is an integer which is greater than another integer**.
- $\forall z (P(z) \wedge E(z))$ translates to **all integers are positive and even**.
- $\forall n (P(n) \Rightarrow E(n))$ translates to **all positive integers are even**.

It is important to note that the translations are typically not literal translations but are translations into grammatically-correct and everyday English. Note also that the translation also abides by the **BNAOIF** precedence rules and thus the determination of the truth or otherwise of these expressions also depends on the precedence rules. For example,

- $\forall x (E(x) \vee \neg E(x))$ translates to **all integers are even or odd**, and is therefore true. However, 34

- $\forall x E(x) \vee \forall y E(y)$ translates to **all integers are even or all integers are odd** which is false.
-

Let the domain be rooms at Kingston Hill, $K(x)$ represent x is in Kenry House, $G(y)$ represent y is on the ground floor and $N(p, q)$ represent p is next to q .

1. Translate each of the following predicate logic expressions into English

- (a) $\exists t K(t)$
- (b) $\forall x (K(x) \wedge G(x))$
- (c) $\forall s (K(s) \Rightarrow G(s))$
- (d) $\forall m (K(m) \vee \neg K(m))$
- (e) $\forall n K(n) \vee \forall x \neg K(x)$
- (f) $\exists m \exists n (K(m) \wedge K(n) \wedge N(m, n))$
- (g) $\forall x \exists y N(x, y)$
- (h) $\exists x \forall y N(x, y)$

2. State whether each expression in question 1 is True or False.

3. Translate each of the following statements into a predicate logic expression.

- (a) All rooms at Kingston Hill are on the ground floor.
- (b) There are two ground floor rooms in Kenry House that are not next to each other.
- (c) There is a room at Kingston Hill which is not on the ground floor.
- (d) All rooms at Kingston Hill are either on the ground floor or not.

4. Let the predicate used in a WHERE clause of an SQL query be called $P(x)$. The domain (as always) is all the values of the fields referred to in the query.

- (a) If $\neg \exists x P(x)$, what is returned by the query.
 - (b) If $\forall y P(y)$, what is returned by the query.
-

5.2 Negating an Expression

The application of the negation operator to a predicate logic expression results in

- each \wedge becoming an \vee ;
- each \vee becoming an \wedge ;
- each \forall becoming an \exists ; and
- each \exists becoming an \forall .

That is,

- $\neg\forall x(P(x) \wedge Q(x))$ is true when **not all are both** true, which is equivalent to **for at least one not both** are true. That is, $\exists x(\neg P(x) \vee \neg Q(x))$; and
- $\neg\exists y(P(y) \vee Q(y))$ is true when **there is not one for which either** are true, which is equivalent to **for all both are not** true. That is, $\forall y(\neg P(y) \wedge \neg Q(y))$.

Using the examples from the previous page:

- $\neg\forall m E(m)$ represents **not all integers are even** which is equivalent to **there is an integer which is odd** written $\exists y \neg E(y)$;
- $\neg\exists n P(n)$ represents **there is not an integer that is positive** which is equivalent to **all integers are odd** written $\forall t \neg E(t)$;

In the following section we present a topic this is essential to the efficient and effective running of any business – **descriptive statistics**.

6 Descriptive Statistics

All businesses have a wealth of data that they need to analyse to review the health of the business and to inform decisions regarding the business' future. In its raw form data is typically impenetrable – just a collection of numbers or values of other types – it needs to be summarised either pictorially or numerically to enable understanding and allow its viewers to draw appropriate conclusions.

There are two categories of statistics:

descriptive statistics which involves summarising data numerically or graphically; and

inferential statistics which involves making inferences about the whole population of data from a sample set of data.

On this module we will only study the first of these categories. The second will be covered on the level 5 module **Business Statistics**.

Now, how one can summarize data – whether one is allowed to use (most) numeric summary measures – depends on the nature of the data.

6.1 Data Categories

There are four categories of data, two of which are **qualitative** – the data has some quality – and two of which are **quantitative** – the data has a quantity and hence is numeric. The qualitative category may be further subdivided into:

Nominal: data which is simply a collection of names. For example, surnames of students or₅₆ addresses of students.

Ordinal: data which is nominal but also has an ordering. That is, one can say that one value is greater than another. For example, ranking of your modules, or module grades. One wouldn't do this with nominal data – one surname cannot be greater than another!

The quantitative category may be further subdivided into:

Interval: data which is ordinal but where the difference between two data values has a consistent meaning. For example, when temperatures are measured in degrees celsius then the difference between two temperatures actually means something. This is certainly not the case with nominal data – one couldn't conceive of subtracting one surname from another – or ordinal data – if A and B were two module grades, what would A - B mean?

Ratio: data which is interval, but where the ratio of two values has a consistent meaning. For example, heights or ages of people. We can say that one person is twice as old as another. We can't do this with interval data – one temperature cannot be twice as large as another – and certainly not with any qualitative data!

Qualitative data, nominal or ordinal, may be summarised by using graphical methods that highlight some quantitative representation of the data, such as the frequency of the values of the data. That is, they may be summarised by using some pictorial representation such as a **frequency distribution**, **bar chart** or a **pie chart**. We can also use one or two of the numerical measures with qualitative data, but not the majority of these measures since they require some arithmetic calculation and this can only be applied to numbers.

Note that although numbers may be used to name or order things these remain nominal or ordinal data. For example, if you rank your modules as 1, 2, 3 etc. these are names of the ranks which have an implicit ordering but are not numbers. You could just as easily have called the ranking values *A*, *B*, *C* etc. or first, second, third and so on.

Classify each of the follows sets of data as nominal, ordinal, interval, or ratio.

- Student surnames;
- Student ID numbers;
- Capacities of teaching rooms;
- Degree classifications;
- Telephone numbers;
- Temperature in fahrenheit;
- Module codes.

6.2 Graphical Methods

Each graphical method presents quantitative measures of a set of data values which may themselves be qualitative or quantitative. That is, they present an (absolute or relative) count of each distinct data value. Each method is presented in the following subsections.

6.2.1 Frequency Distribution

A frequency distribution does exactly what it says on the tin! It presents in tabular form the **frequency** of each data value. For example, the genders of the students on this module – nominal data – may be summarised in the following frequency distribution.

Data Value	Frequency
Male	74
Female	32

Quantitative data may also be summarised using graphical methods. Here is a (not completely accurate) frequency distribution of the ages of students on this module.

Data Value	Frequency
18	55
19	29
20	15
21	5
22	1
23	1

Note that the sum of the frequencies must equal the total number of students on the module.

6.2.2 Relative Frequency Distribution

A relative frequency distribution also does exactly what it says on the tin! It presents in tabular form the **relative frequency** of each data value. In this table instead of presenting frequencies, we present frequencies divided by the total number of data. For example, the genders of the students on this module – nominal data – may be summarised in the following relative frequency distribution.

Data Value	Relative Frequency
Male	$\frac{74}{106}$
Female	$\frac{32}{106}$

Note that the relative frequencies may be presented as decimals, but since the sum of the relative frequencies must be 1, some rounding may be required.

Here is a (not completely accurate) relative frequency distribution of the ages of students on this module.

Data Value	Relative Frequency
18	0.52
19	0.27
20	0.14
21	0.05
22	0.01
23	0.01

Clearly a relative frequency distribution may be generated directly from a frequency distribution by simply dividing each frequency by the total number of data values.

6.2.3 Percentage Frequency Distribution

A percentage frequency distribution also does exactly what it says on the tin! It presents in tabular form the **percentage frequency** of each data value. In this table instead of presenting relative frequencies, we present frequencies as a percentage of the total number of data. For example, the genders of the students on this module – nominal data – may be summarised in the following relative frequency distribution.

Data Value	Percentage Frequency
Male	69.8%
Female	30.2%

Note that the percentage frequencies must sum to 100 and may be generated from a relative frequency distribution by simply multiplying each relative frequency by 100.

Here is a (not completely accurate) percentage frequency distribution of the ages of students on this module.

Data Value	Percentage Frequency
18	52%
19	27%
20	14%
21	5%
22	1%
23	1%

6.2.4 Cumulative Distributions

For ordinal, interval and ratio data we may also produce **cumulative** distributions which present the data values alongside the frequency of that value or less. Note that for all of our distributions we have presented the data values in ascending order, and thus, the cumulative frequencies can be generated by simply summing frequencies from the non-cumulative distributions.

For example, the cumulative frequency distribution and cumulative relative frequency distribution for the age data are presented below.

Data Value	Cumulative Frequency
18	55
19	84
20	99
21	104
22	105
23	106

Data Value	Cumulative Relative Frequency
18	0.52
19	0.79
20	0.93
21	0.98
22	0.99
23	1

1. Produce frequency, relative frequency, and percentage frequency distributions for the following set of data which presents (the made-up) nationalities of students on this module.

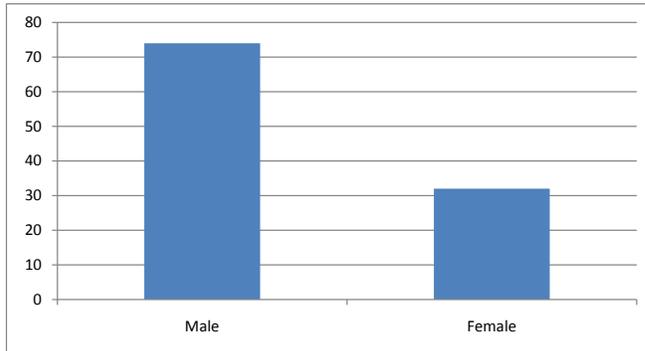
British American French British Polish Hungarian British
Iranian Polish British Spanish British British Indian
British Latvian British Pakistani American British British
Italian Chinese British British British Indian Afghan
British British British American French British Polish
Hungarian British Iranian Polish British Spanish British
British Indian British Latvian British Pakistani American
British British Italian Chinese British British British
Indian Croatian British British Italian Brazilian French
British Polish Hungarian British Iranian Polish British
Spanish British British Indian British Polish British
Pakistani American British British Italian Chinese Spanish
British British Indian Afghan British British British
American French British Polish Mexican British Iranian
Polish British Spanish British Croatian Indian British Dutch

2. Produce frequency, relative frequency, percentage and cumulative frequency distributions for the following set of data which presents (the made-up) number of siblings of students on this module.

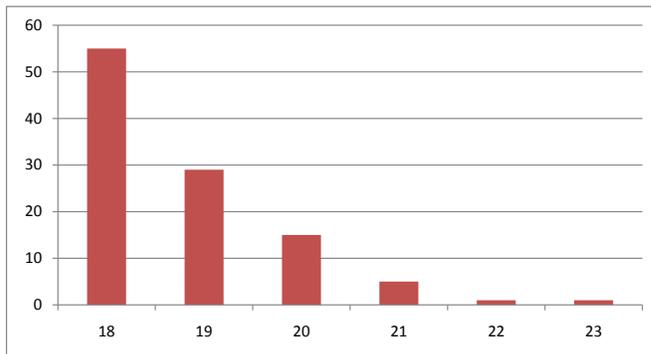
0 0 1 0 3 4 2 2 1 0 0 3 4 1 1 1 1 2 1 2 0 3 4 2 0 1 0 1 1 2 0 0 0 0
3 4 2 2 1 0 0 3 4 1 5 1 1 2 1 2 0 3 4 2 0 1 0 2 1 2 0 0 1 0 1 4 1 2
1 0 0 3 4 0 1 1 1 2 1 2 0 3 4 2 0 1 0 1 1 2 0 0 0 0 3 4 1 2 1 0 0 3
0 1 3 1

6.2.5 Bar Chart

Sometimes a picture is preferable to a table and more appropriately presents the details of the summarised data. A frequency distribution may be presented in a **bar chart** which simply presents each frequency as a bar whose size reflects the size of the frequency. The bars may be presented vertically, horizontally, in two dimensions or even in three dimensions. Here is a two dimensional, vertical bar chart of the genders of students on the module.



Here is a bar chart of the (not completely accurate) frequency distribution of the ages of students on this module.

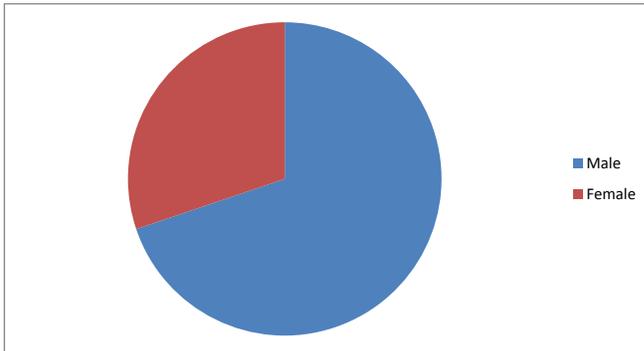


The shape of the distribution is clearer on a bar chart than in a frequency distribution. For example, the age data bar chart clearly indicates that the data is **right-skewed** which means that it tails off to the right. The other alternatives are **left-skewed** (tails off to the left) and **symmetric** which mean that it neither tails off to the left nor right but is symmetric about some central value.

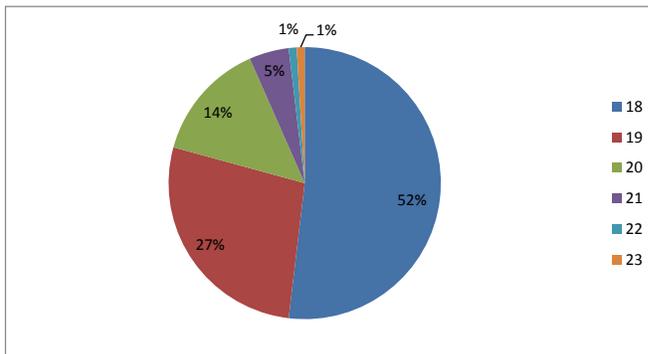
6.2.6 Pie Chart

A frequency distribution may also be presented in a **pie chart** which simply presents each frequency as a segment of a circle (a pie) whose size reflects the size of the frequency. The pie chart may be presented in two or three dimensions. Interestingly Florence Nightingale – who has links to Kingston University – is credited with developing a form of a pie chart.

Here is a pie chart for the gender data.



Here is a pie chart of the (not completely accurate) frequency distribution of the ages of students on this module.



Produce a bar chart and pie chart for the nationality and sibling data.

We now present the numerical methods for summarising (typically) quantitative data. They are presented within two categories:

- **measures of central tendency** which describe the ‘middle’ of the data; and
- **measures of dispersion** which describe how the data is spread.

The measures of central tendency are the mode, median and mean.

6.2.7 Mode

The **mode** of a collection of data is the **most frequent** value or values. Since it is simply a frequency count and does not involve any arithmetic manipulation of data values, it may be applied to any data, nominal, ordinal, interval, or ratio. For our student gender data the mode is **male**, and for the age data the mode is **18**.

6.2.8 Median

The **median** is the **middle value of the data**. That is, if the data were presented in ascending order then the median would be the middle one. Of course, if there is an even number of data then there isn't a middle one – the median is the mean of the middle two. Now, at most 50% of the data is less than the median and at most 50% of the data is greater than the median. The median is also known as the **50th percentile** and the **2nd quartile**. We discuss percentiles and quartiles in section 6.2.11.

Since nominal data does not have an order there is no way they can be placed in ascending order. Thus nominal data cannot have a median. Although ordinal data may have a median there are problems if there is an even number of data since we cannot generate a mean of ordinal data.

We cannot therefore calculate a median for the gender data. The median of the age data is 18. Note that this can be determined directly from the frequency distribution which presents the data in ascending order. The middle two values – there is an even number of data – are the 53rd and 54th values, both of which are in the first group and thus the median is 18.

6.2.9 Mean

The mean of a set of data is typically calculated in one of two ways:

- sum the data and divide by the number of values; or
- the sum of the product of values and their relative frequency.

Thus for the age data we can generate the mean by summing the 106 values and then dividing by 106. That is,

$$\frac{18 + 18 + \dots 18 + 19 + \dots + 19 + \dots 22 + 23}{106} = 18.78$$

We could have generated this by summing the products of the data values and their frequencies.

The second approach directly used the relative frequency distribution. We add an extra column to present the products of the data values and their relative frequencies. Then simply sum these to generate the mean. That is,

Data Value	Relative Frequency	Data Value * Relative Frequency
18	0.52	9.36
19	0.27	5.13
20	0.14	2.80
21	0.05	1.05
22	0.01	0.22
23	0.01	0.23
Sum		18.79

Note that the two means are not exactly the same. That is because of the rounding of the relative frequencies to two decimal points.

When analyzing a set of (quantitative) data it is worthwhile calculating all three measures of central tendency since we can draw conclusions about the distribution of the data from these values. For example,

- if the mode, median and mean are the same value then the distribution is **symmetric** about a central peak;
- if the mean is greater than the median, and the median is greater than the mode, then the data is **right-skewed**; and
- if the mean is less than the median, and the median is less than the mode, then the data is **left-skewed**.

1. Generate the mode, median, and mean for the sibling data.
2. Is the distribution symmetric, right-skewed or left-skewed?
3. (a) If fourteen more students joined the module with 3, 1, 1, 2, 0, 0, 1, 1, 0, 2, 1, 1, 2 and 1 sibling respectively, how would this affect the mode, median and mean?
(b) How would these new students affect the shape of the distribution?

In the following sections we present the measures of dispersion which provide an indication of the variation of the data. We begin with the range.

6.2.10 Range

The range of a set of data is simply the difference between the largest and smallest value. Thus for the age data the range is 5 years. Clearly the range may only be calculated for quantitative data.

6.2.11 Inter-Quartile Range

This measure is dependent on two other values – the first and third quartile. Note that the

- **first quartile**, also known as the **25th percentile**, is the smallest value that is greater than or equal to 25% of the data; and the
- **third quartile**, also known as the **75th percentile**, is the smallest value that is greater than or equal to 75% of the data, or equivalently, is the largest value that is less than or equal to 25% of the data.

The inter-quartile range is the difference between the 3rd quartile and the 1st quartile. It indicates the variation of the middle 50% of the data. It thus ignores possible extreme values (or outliers) that impact on the calculation of the range.

The calculation of the quartiles (and percentiles in general) proceeds as follows.

If there are n data values sorted in ascending order then the position of the P th percentile is calculated as follows:

$$PV = n * \frac{P}{100}$$

Now if

- PV is an integer, then the value is the mean of those in positions PV and PV + 1; and if
- PV is not an integer then round it up to the nearest integer.

Note that this algorithm also works when calculating a median, which is the **2nd quartile** or **50th percentile**.

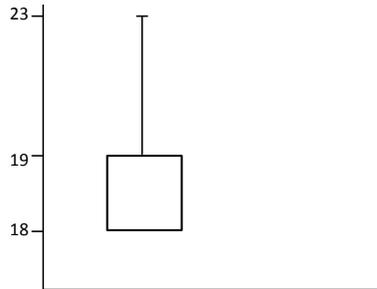
Thus for the age data the 25th percentile is found in position 27 (26.5 rounded up to 27), and the 75th percentile is found in position 80 (79.5 rounded up to 80). The values in these positions are 18 and 19 respectively. Thus the inter-quartile range is 1 year.

It is clear that percentiles, quartiles, and the inter-quartile range, may only be calculated for quantitative data.

As with frequency distributions we may produce a pictorial representation of the range, inter-quartile range and median. They are represented on a **box plot** or **box-and-whisker diagram**. This is simply a

- box which highlights the middle 50% of the data – it is bordered by the first quartile and third quartile;
- whiskers emanating from the box whose limits on the maximum and minimum data values – the distance between them is the range; and
- the median which appears somewhere in the box, since the median or second quartile must be greater than or equal to the first quartile and less than or equal to the third quartile.

We present the box plot for the age data below. Note that the median is equal to the first quartile and there is no lower whisker since the minimum value is equal to the first quartile.



-
1. Create a box plot for the sibling data.
 2. What conclusions can you draw from the box plot?
-

Note that in American Universities and schools it is typical to provide both a grade for a module and a percentile figure which indicates a student's performance alongside his or her peers. For example, if a student's mark is at the 90th percentile then this student is in the top 10% of the class.

6.2.12 Variance and Standard Deviation

The variance and standard deviation are two related and very important measures of dispersion that are ubiquitous throughout the application of statistics. Since the standard deviation is simply the square root of the variance we only need to explain the calculation of the variance.

The variance is the **mean squared difference from the mean**. That is, it measures how much the data is spread about the mean. To generate the variance one simply:

- subtracts the mean from each data value – the difference from the mean;
- square all of the differences; and
- then calculate the mean of these squares.

We square the differences because we are only interested in how far away a value is from the mean and not if it is less than or greater than the mean. If the differences were not squared then negative and positive differences would cancel each out, and a data set which is significantly

spread about the mean would not be indicated by this measure. In actual fact, if we didn't square the differences then the result would always be zero.

The variance is best calculated using a frequency distribution. Since a variance may only be calculated for quantitative data we will generate the variance for the age data.

Data Value	Frequency	Diff. from the Mean	Square Diff
18	55	-0.79	0.624
19	29	0.21	0.04
20	15	1.21	1.464
21	5	2.21	4.884
22	1	3.21	10.304
23	1	4.21	17.724
Variance			1.037

Note that the variance is generated by multiplying the squared differences by the frequencies, summing these values and then dividing by 106 – that is the mean of the squared differences.

Most textbooks present a formula for generating a mean that involves the summation sign \sum which simply means add up a collection of values. For example,

$$\sum_{i=1}^3 x_i = x_1 + x_2 + x_3$$

and

$$\sum_{j=0}^4 2^j = 2^0 + 2^1 + 2^2 + 2^3 + 2^4 = 31$$

Now the variance formula is

$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

where μ represents the mean and x_i name each of the data values x_1, x_2 and so on. That is, we add up a bunch of squared differences and divide by the number of data values – the mean of the squared differences!

It is clear from the way we generate the variance that one can never have a negative variance – since we are dealing with squares. It is also clear that values that are a long way from the mean are emphasized by the variance calculation, since we square the differences.

The **standard deviation** of the age data is $\sqrt{1.037} = 1.018$.

6.2.13 Z-score

Now we can use the standard deviation to determine how far a value is away from the mean relative to the size of the standard deviation. That is, we can produce a measure, called the **Z-score or standard score**, in units of the standard deviation that allows us to compare data values from different data sets. We generate this value by calculating the difference from the mean and then dividing by the standard deviation. Thus an age of 23 is

$$\frac{23 - 18.79}{1.018} = \frac{4.21}{1.018} = 4.14$$

4.14 standard deviations away from the mean.

Now the vast majority (89%) of data lies within 3 standard deviation either side of the mean. Any value beyond 3 standard deviations away from the mean is regarded as an **outlier**

6.2.14 Comparing Data Sets

We can (and often do) use the variance in combination with the mean to compare different sets of data. For example, we may want to compare the student performance on different modules. Lets assume that:

- module A has a mean mark of 53 and a standard deviation of 4;
- module B has a mean mark of 58 and a standard deviation of 7; and
- module C has a mean mark of 60 and a standard deviation of 2.

We may draw the following conclusions about the modules:

- the assessment on module C is the least discriminating or the students on the module are the most homogeneous. This is because the data has the least spread – about 90% of the data lies between marks of 54 and 66;
- the assessment on module B is the most discriminating or the students on the module are the least homogeneous. This is because the data has the most spread – about 90% of the data lies between marks of 37 and 79; and
- the assessment on module A is somewhere in the middle. About 90% of the data lies between marks of 41 and 65.

To add to our understanding of the data we could also generate a box plot. In summary, the more measures available to the analyst the better the conclusions that will be reached.

-
1. Generate the variance and standard deviation for the age data.
 2. Does the age data include any outliers?
 3. The students on another module have a mean age of 19.3 and a variance of 1.7. What conclusions can you draw about the ages of the students on the two modules?
 4. What is the minimum value for a variance and what can you say about data that have this minimum variance?
 5. What is the maximum variance value for a particular data set?
 6. Prove that if you don't square the differences then the measure will always equal 0.
-

7 Probability Theory

Most students on this module will have some prior experience of probability. We often make claims about something having no chance of happening or something being certain to happen. Whether one takes part in the National Lottery, are paying car insurance or are producing predictive SATs results, probability plays an essential role. That is,

- the National Lottery payouts are based on the likelihood (probability) of getting a certain number of the lottery numbers correct;
- your car insurance premium is based on the likelihood (probability) of you having an accident; and
- predicting a child's SATs scores is calculated using the likelihood (probability) of achieving various SATs scores.

Probability theory is used in all businesses.

- A university has to plan for future student numbers based on the likelihood of a student completing a programme, and the likelihood of an applicant attending the university.
- A manufacturing company will want to minimize the (potential) cost of the introduction of new products by trying to predict the chance of success.
- A financial services company will base most of their decisions on the likelihood of making a return on any investment.

In the following section we introduce probability and describe how it is dependent on **set theory** and **functions**.

7.1 Probability

Defining the term probability is not as easy as one may think. Wikipedia provides the following insight:

The word probability does not have a consistent direct definition. In fact, there are sixteen broad categories of probability interpretations, whose adherents possess different (and sometimes conflicting) views about the fundamental nature of probability:

- **Frequentists** talk about probabilities only when dealing with experiments that are random and well-defined. The probability of a random event denotes the relative frequency of occurrence of an experiment's outcome, when repeating the experiment. Frequentists consider probability to be the relative frequency "in the long run" of outcomes.
- **Bayesians**, however, assign probabilities to any statement whatsoever, even when no random process is involved. Probability, for a Bayesian, is a way to represent an individual's degree of belief in a statement, given the evidence.

(Wikipedia, <http://en.wikipedia.org/wiki/Probability>)

Although later in this section we will discuss examples that use the Bayesian approach, for now we will define probability using the frequentist approach.

We will animate our definition by using an example which will be revisited throughout this section. We start with the set of students on this module. We call this set the **sample space** which is normally denoted Ω . In common with the universal set from set theory, this is the set of all possible values for a particular problem. We can define the sample space Ω as

$$\Omega = \{x_1, x_2, \dots, x_n\}$$

where x_1 etc. are the names of the elements of the sample space, in this case the students on this module.

We now need to define a very important function – the **probability function** f – which determines how we calculate probabilities.

Definition 25 Probability Function The probability function f has:

- domain the sample space,
- codomain the real numbers between (and including) 0 and 1, and
- $f(x_1) + f(x_2) + \dots + f(x_n) = 1$. That is, the sum of the probabilities of all of the values in the sample space is 1 which is otherwise denoted as $\sum_{i=1}^n f(x_i) = 1$.

Thus a valid probability function simply maps the elements of the sample space – the possible outcomes of a particular experiment – to a value between 0 and 1, and the sum of the application of the function to each input is 1. Immediately we can draw the conclusions that:

- **all probabilities lie between 0 and 1**. That is, there are no negative probabilities and no probabilities greater than 1! This is consistent with viewing probabilities as relative frequencies, and

- the sample space includes all possible outcomes of a particular experiment since the sum of the probabilities of its elements is 1. Note that we previously learnt in section 6.2.2 that relative frequencies sum to 1.

For many problems, the probability of each element in the sample space being selected is the same. That is, the relative frequency of each element is the same – they are all **equally likely**. For example,

- the probability of selecting any student on this module is $\frac{1}{106}$ (1 divided by the number of students on this module);
- the probability of getting any number on a fair dice is $\frac{1}{6}$;
- the probability of selecting any number first in the National Lottery is $\frac{1}{49}$ – this of course changes for the second selection and so on; and
- the probability of selecting any card from a pack of cards is $\frac{1}{52}$.

In our examples above we are looking at the probabilities of selecting one particular element from the sample space, but in most applications of probability we are interested in the likelihood (probability) of selecting **any of a particular set of elements**. That is, selecting any from a **particular subset of the sample space**. For example, the probability of selecting an 18 year old student or the probability of selecting a male student. We call these subsets of the sample space **events**.

Definition 26 Probability of an Event The probability of an event E is written $P(E)$. Now

$$P(E) = \sum_{x_i \in E} f(x_i)$$

That is, **the probability of an event is the sum of the probabilities of the elements of the event**.

Now

- if $E = \emptyset$ then $P(E) = 0$ which implies that E is impossible; and
- if $E = \Omega$ then $P(E) = 1$ which implies that the event is certain.

If we let ES be the event eighteen year old student then $ES \subseteq \Omega$ and

$$P(ES) = \sum_{x_i \in ES} f(x_i) = 55 * \frac{1}{106} = \frac{55}{106} = 0.52$$

That is, since each student is equally likely to be selected, the probability of selecting any of the 55 eighteen year old students is **the probability of selecting one of them multiplied by 55**. Note that 0.52 is simply the relative frequency of eighteen year old students as presented in the relative frequency distribution in section 6.2.2.

Similarly if we let M represent the event male student then

$$P(M) = \frac{74}{106} = 0.7$$

-
- What is the probability of selecting a female student?
 - What is the probability of getting an even number on a dice?
 - What is the probability of selecting a red card from a pack of cards?
 - If an event has a probability of 0.9 what can you say about the event?
 - If an event has a probability of 0.5 what can you say about the event?
 - Why is $P(\Omega) = 1$?
-

We now present some of the probability rules which are all directly linked to the application of one or more set operators.

7.2 Probability Rules

We begin with the complement rule.

7.2.1 Complement Rule

For any event $E \subseteq \Omega$

$$P(E^c) = 1 - P(E)$$

That is, for any event **the probability of its complement is 1 minus the probability of the event.**

This rule can be quite easily derived. We know from section 2 that $E^c = \Omega \setminus E$ which in probability terms is rewritten as $E^c = \Omega \setminus E$. Thus

$$P(E^c) = P(\Omega \setminus E)$$

Now we also know that the set difference operator just removes elements from the left hand set that are also in the right hand set. In this case since all elements of E are in the sample space, every element of E is removed. Thus

$$P(\Omega \setminus E) = P(\Omega) - P(E)$$

and since $P(\Omega) = 1$, we get

$$P(\Omega \setminus E) = 1 - P(E)$$

- If an event is impossible, what can you say about its complement?
- Can an event have the same likelihood as its complement?
- For two events, A and B , if $P(A) + P(B) = 1$ does this mean that A and B are complementary events?

7.2.2 Joint Probability

Joint probability refers to any situation where one wants to know the probability of two events occurring at the same time. For example, what is the probability of selecting an eighteen year old, male student. Assuming that there are 37 eighteen year old, male students, the probability is $\frac{37}{106} = 0.35$.

In general, if A and B are two events then

$$P(A \cap B)$$

represents the joint probability of A and B . Now,

- if $A \cap B = \emptyset$ then $P(A \cap B) = 0$. That is, if A and B have nothing in common – are **mutually exclusive** – then it is impossible for them both to occur at the same time;
- if $A \subseteq B$ then $P(A \cap B) = P(A)$ since $A \cap B = A$; and
- if A and B are **statistically independent** then $P(A \cap B) = P(A) * P(B)$.

The first and third items in the list require two new definitions.

Definition 27 Mutually Exclusive Events Two events are mutually exclusive (or disjoint) if they have no elements in common. That is, if one event occurs then the other cannot occur. For example, the events **even number** and **odd number** are mutually exclusive, as are **18 year old student** and **19 year old student**. If two events A and B are mutually exclusive then $A \cap B = \emptyset$. Thus, an event and its complement are always mutually exclusive.

Definition 28 Statistically Independent Events Two events are statistically independent if the occurrence of one does not affect the likelihood of the occurrence of the other. For example,

- if the sample space is the integers, then the events **positive integer** and **even integer** are statistically independent, since the likelihood of selecting an even number is not affected by knowing that the number is positive;
- if the sample space is the letters of the alphabet (upper and lower case), the events **lower case letter** and **vowel** are statistically independent, since the likelihood of selecting a vowel is not affected by knowing that the letter is lower case.

In both of these examples, the proportion of elements of the sample space that are in the second event is the same as the proportion of elements of the first event that are in the second event. That is,

- half of the integers are even and half of the positive integers are even;
- $\frac{5}{26}$ ths of the (upper and lower case) letters are vowels and $\frac{5}{26}$ ths of the lower case letters are vowels.

Hence, the **probability of selecting an even, positive integer** is simply the product of the **probability of selecting an even integer** and the **probability of selecting a positive integer**. That is, if Ev represents even integer and Po represents positive integer then

$$P(Ev \cap Po) = P(Ev) * P(Po) = 0.5 * 0.5 = 0.25$$

As another example, if we let the sample space be a pack of cards, D represent diamond card and Pi represent picture card (jack, queen, king) then,

$$P(D \cap Pi) = P(D) * P(Pi) = \frac{1}{4} * \frac{3}{13} = \frac{3}{52}$$

Of course, most events are not statistically independent. That is, the occurrence of one event does affect the likelihood of the occurrence of the other. In these cases, we cannot use the rule presented above but instead have to count the number of elements in the intersection of the two events. For example,

- if the sample space is a pack of cards then the events red card (denoted R) and diamond card are not statistically independent. That is,

$$P(R \cap D) = 0.25 \neq P(R) * P(D) = 0.5 * 0.25 = 0.125$$

- if the sample space is the students on this module then the events male student and female student are not statistically independent. That is,

$$P(M \cap M^c) = 0 \neq P(M) * P(M^c) = 0.7 * 0.3 = 0.21$$

Clearly, the occurrence of the event male student will affect the likelihood of the occurrence of the event female student!

1. Which of the following pairs of events are statistically independent.
 - even number and number between 1 and 10 (sample space is the integers);
 - odd number on a dice and prime number on a dice (sample space is the numbers on a dice);
 - vowel and letters from A to M (sample space is the letters of the alphabet);
 - vowels and consonants (sample space is the letters of the alphabet);
 - male students and eighteen year old students (sample space is students on the module);
 - male students and students over 6 feet tall (sample space is students on this module);
 - male students and students over 35 years old (sample space is students on this module);
 - numbers between 1 and 4 and prime numbers (sample space is the integers between 1 and 8)
 2. Which of the pairs of events are mutually exclusive?
 3. Evaluate the joint probabilities of each pair of events.
 4. If $P(A) = 0.7$, $P(A \cap B) = 0.5$ what is $P(A \cap B^c)$?
-

7.2.3 Addition Rule

We are often interested in the likelihood of either of two events occurring such as selecting a male or eighteen year old student. Now one would assume that the **probability of selecting a male or eighteen year old student** written $P(M \cup ES)$ would be

$$P(M \cup ES) = P(M) + P(ES) = 0.52 + 0.7 = 1.22$$

But we have ended up with a probability that is greater than 1!

This is because we have counted certain elements twice. The event M includes eighteen year old students, and the event ES includes male students. That is, the union of two sets results in the set that contains the elements that are in **at least one of the sets** which includes the possibility that elements are in both sets. Thus for $A \cup B$ elements may

- only be in A ;
- only be in B ; or
- be in both A and B

If we simply sum the probability of each event we would count those elements that are in both events twice! Thus we need to subtract the probability of being in both sets. In our example we need to subtract the probability that a student is both eighteen years old and male. Thus we get,

$$P(M \cup ES) = P(M) + P(ES) - P(M \cap ES) = 0.52 + 0.7 - 0.35 = 0.87$$

In general,

$$P(A \cup B) = \sum_{x \in (A \cup B)} f(x) = \sum_{x \in A} f(x) + \sum_{x \in B} f(x) - \sum_{x \in (A \cap B)} f(x)$$

and hence, for any two events A and B , the likelihood of selecting one or the other is represented using the following addition rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

We then have two special cases to deal with.

- if A and B are **mutually exclusive** then $P(A \cup B) = P(A) + P(B)$, we simply sum the probabilities since the events have no elements in common and thus nothing is counted twice; and
- if A and B are **statistically independent** then $P(A \cup B) = P(A) + P(B) - P(A) * P(B)$

We will end this section with one further definition.

Definition 29 Collectively Exhaustive Events Two or more events are collectively exhaustive if their union is equal to the sample space. Thus, in the case of two events, say A and B , they are collectively exhaustive if

$$A \cup B = \Omega$$

Alternatively we can represent this using probability by stating that A and B are collectively exhaustive if

$$P(A \cup B) = 1$$

1. Evaluate

(a) $P(ES \cup M^c)$

(b) $P(NS \cup M)$ assuming that NS represents nineteen year old students and that there are 20 male nineteen year old students.

(c) $P(ES \cup NS)$

(d) $P(M \cup M^c)$

2. If $P(A) + P(B) > 1$ what can you say about the events A and B .

3. If $P(A) = p$ and $P(A \cup B) = q$, what is $P(A \cup B^c)$?

4. If two events are mutually exclusive and collectively exhaustive, what can you say about the events?

7.2.4 Conditional Probability

Conditional probabilities are an essential tool when making business decisions using market research data. They enable the updating of a likelihood due to new information – say, the results of market research or some related experience.

Now we have previously calculated the probability of selecting an eighteen year old, male student, and the probability of an eighteen year old or male student. But what about the probability of selecting an eighteen year old given that the student is male. This may sound the same as selecting an eighteen year old, male student but is different because

- the probability of selecting an eighteen year old, male student is the likelihood of selecting someone with both of these characteristics from the whole set of students; while
- the probability of selecting an eighteen year old given that the student is male, is the likelihood of selecting an eighteen year old from the subset of male students.

Definition 30 Conditional Probability The conditional probability of the event A given the event B , is the probability of A occurring given that B has occurred. It is written $P(A|B)$. In general

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

That is, **the probability of A given B is equal to the probability of A and B divided by the probability of B.**

Returning to our example, $P(ES|M)$ is the probability of selecting an eighteen year old student given the student is male. That is, we are now restricting the selection space to the male students, and then want to select an eighteen year old. In other words, what proportion of the male students are also eighteen years old. To generate this probability we need to know how many students are male and how many students are male and eighteen.

Since there are 74 male students and 37 eighteen year old, male students,

$$P(ES|M) = \frac{37}{74} = 0.5$$

Now this is equivalent to evaluating this by dividing $P(ES \cap M)$ by $P(M)$. That is,

$$P(ES|M) = \frac{P(ES \cap M)}{P(M)} = \frac{\frac{37}{106}}{\frac{74}{106}} = \frac{37}{74} = 0.5$$

Interestingly normal probability is a special case of conditional probability where the given event is the sample space S . Thus,

$$P(A|S) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1} = P(A)$$

since the occurrence of the sample space is certain and thus imposes no real condition on the other event.

We end this section in the same way that we ended the previous section by focusing on two special cases – mutually exclusive events and statistically independent events. Now if,

- A and B are mutually exclusive events then $P(A|B) = 0$ since the occurrence of one event precludes the occurrence of the other; and if
- A and B are statistically independent and $P(B) \neq 0$, then $P(A|B) = P(A)$ since the occurrence of one does not effect the likelihood of the other. This is derived as follows

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A)$$

1. Evaluate

- $P(ES|M^c)$
- $P(ES|NS)$
- $P(D|Pi)$
- $P(Pi|D)$

2. For any event A what is $P(A|A^c)$?

3. If for two events A and B , $P(A|B) = P(B|A)$, what can you say about the events A and B .

7.3 Bayes' Theorem

Bayes' theorem describes a conditional probability in terms of its inverse. That is, if we have two events A and B we can define $P(A|B)$ by using its inverse $P(B|A)$. The relationship is

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

It is named after the Reverend Thomas Bayes. We will jump straight into this with an illustrative example.

We know that $P(ES|M) = 0.5$, but we will now show how this may be calculated using the inverse conditional probability $P(M|ES)$. Using Bayes' theorem we get

$$P(ES|M) = \frac{P(M|ES) * P(ES)}{P(M)}$$

Now $P(ES) = \frac{55}{106}$, $P(M) = \frac{74}{106}$, and $P(M|ES) = \frac{37}{55}$. Thus,

$$P(ES|M) = \frac{\frac{37}{55} * \frac{55}{106}}{\frac{74}{106}} = \frac{\frac{37}{106}}{\frac{74}{106}} = \frac{37}{74} = 0.5$$

Thus we know it works but

- why does it work?
- and why is it important?

We will answer these questions in the following subsections.

7.3.1 Why does it work?

We know from the definition of conditional probability that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Thus in our example we have

$$P(M|ES) = \frac{P(M \cap ES)}{P(ES)}$$

and when we substitute $P(M|ES)$ by $\frac{P(M \cap ES)}{P(ES)}$ in the Bayes Theorem expression

$$P(ES|M) = \frac{P(M|ES) * P(ES)}{P(M)}$$

we get

$$P(ES|M) = \frac{\frac{P(M \cap ES)}{P(ES)} * P(ES)}{P(M)}$$

which simplifies to

$$P(ES|M) = \frac{P(M \cap ES)}{P(M)}$$

which is the definition of conditional probability!

1. Using Bayes' theorem evaluate

(a) $P(M|ES)$

(b) $P(ES|F)$

2. Show how Bayes' theorem works for two mutually exclusive events.

7.4 Why is it important?

There are many situations where we need to calculate one conditional probability using its inverse. A classic example – and a business example – is when we want to update the likelihood of success of a new product based on market research results. We will illustrate this with an example from close to home.

Kingston Business School is considering introducing a new Masters programme. The development of a new programme is expensive, and even more expensive if the programme fails to recruit students. It is therefore in the Business School's interest to have as much information as possible regarding the likelihood of success of the new programme. Based on previous experience with new programmes they believe that the new programme has a probability of success of 0.6 (and thus the probability of failure is 0.4). Note that these are not **frequentist** probabilities since they are not based on relative frequencies, but are **Bayesian** probabilities based on people's beliefs.

A member of the Department of Strategy, Marketing and Entrepreneurship has recommended that the Business School should make use of a market research company that have some expertise in education-based research. For a known fee the company will carry out research which will predict the likelihood of success of the new programme. They claim that their research predicts success correctly 90% of the time, and it also predicts failure correctly 80% of the time.

The Business School will use this information in association with financial information regarding the cost of the market research, and financial implications of the programme being a success or failure, to decide on its course of action.

The first thing that we need to do is to identify – and name – the events described in this problem. They are:

- S which represents the event the programme is a success;
- MRS which represents the event the market research predicts success.

The next thing to do is to represent the probabilities listed in the problem using these named events.

$$P(S) = 0.6 \quad (7.1)$$

$$P(S^c) = 0.4 \quad (7.2)$$

$$P(MRS|S) = 0.9 \quad (7.3)$$

$$P(MRS^c|S) = 0.1 \quad (7.4)$$

$$P(MRS^c|S^c) = 0.8 \quad (7.5)$$

$$P(MRS|S^c) = 0.2 \quad (7.6)$$

Now,

- 7.1 and 7.2 represent the probabilities that the programme is a success and failure respectively. These are regarded as **prior probabilities** since they are estimated in advance of any further information;
- 7.3 to 7.6 are conditional probabilities representing the predictive reliability of the market research;
 - 7.3 – the probability of the market research predicting success given a successful programme – represents the reliability of a successful prediction, and thus
 - 7.4 represents the non-reliability of a successful prediction;
 - 7.5 – the probability of the market research predicting failure given a failed programme – represents the reliability of a failure prediction, and thus
 - 7.6 represents the non-reliability of a failure prediction.

Now in considering whether to use the market research company, the Business School needs to determine the likelihood of success of the new programme given that the market research predicts success. That is, $P(S|MRS)$. This is known as a **posterior probability** – as opposed to a prior probability – since its value is based on new information. It is the inverse of a known conditional probability $P(MRS|S)$ (the reliability of a successful prediction), and therefore, we can use Bayes' Theorem to generate its value. That is,

$$P(S|MRS) = \frac{P(MRS|S) * P(S)}{P(MRS)}$$

Unfortunately we do not have a value for $P(MRS)$. However, we know that for any two events A and B

$$A = (A \cap B) \cup (A \cap B^c)$$

and thus

$$P(A) = P((A \cap B) \cup (A \cap B^c))$$

and by using the addition rule of probability we get

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

since the intersection of $A \cap B$ and $A \cap B^c$ is the empty set.

Thus if we use the events MRS and S we get,

$$P(MRS) = P(MRS \cap S) + P(MRS \cap S^c)$$

Now since

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

by algebraic manipulation we get

$$P(A \cap B) = P(A|B) * P(B)$$

Thus by substituting

- $P(MRS \cap S)$ with $P(MRS|S) * P(S)$; and
- $P(MRS \cap S^c)$ with $P(MRS|S^c) * P(S^c)$

we get

$$P(MRS) = P(MRS|S) * P(S) + P(MRS|S^c) * P(S^c)$$

Hence, the Bayes' Theorem expression

$$P(S|MRS) = \frac{P(MRS|S) * P(S)}{P(MRS)}$$

becomes

$$P(S|MRS) = \frac{P(MRS|S) * P(S)}{P(MRS|S) * P(S) + P(MRS|S^c) * P(S^c)}$$

Now we have all of the values of these probabilities – presented in equations 7.1 to 7.6 – and can therefore simply substitute them into the expression.

Thus

$$\begin{aligned} P(S|MRS) &= \frac{0.9 * 0.6}{0.9 * 0.6 + 0.2 * 0.4} \\ &= \frac{0.54}{0.62} \\ &= 0.87 \end{aligned}$$

So if the Business School decides to use the market research company and they carry out the market research and determine that the new programme will be a success, then the probability of it actually being a success if launched is 0.87. This is a significant increase from the prior probability of 0.6.

Of course, we should also calculate the likelihood of success if the market research suggests failure. That is,

$$P(S|MRS^c) = \frac{P(MRS^c|S) * P(S)}{P(MRS^c|S) * P(S) + P(MRS^c|S^c) * P(S^c)}$$

Thus,

$$\begin{aligned} P(S|MRS^c) &= \frac{0.1 * 0.6}{0.1 * 0.6 + 0.8 * 0.4} \\ &= \frac{0.06}{0.38} \\ &= 0.16 \end{aligned}$$

So if the Business School decides to use the market research company and they carry out the market research and determine that the new programme will be a failure, then the probability of it actually being a success if launched is 0.16. This is a significant decrease from the prior probability of 0.4.

The Business School can now use this extra information with the relevant financial implications of success or failure, along with the cost of using the market research, to determine their course of action. Should they use the market research, and if so, should they launch the new programme? They would typically use a pictorial method known as a **decision tree** as an aid in this decision-making process.

1. Use Bayes' theorem to confirm that

(a) $P(S^c|MRS) = 0.13$

(b) $P(S^c|MRS^c) = 0.84$

2. What are $P(MRS)$ and $P(MRS \cap S)$?

In the following and final section of this document we present a collection of **discrete probability distributions**. These distributions can be used to determine probabilities of events in everyday problems.

8 Discrete Probability Distributions

In the previous section we studied probability theory and how to calculate various probabilities such as joint and conditional probabilities. The probabilities associated with many problems can be modelled by known probability distributions. These exist in one of two categories:

- **discrete probability distributions** where the values for which we determine probabilities are **discrete** such as the face value of a dice; or
- **continuous probability distributions** where we determine the probability of being within a **continuous** range of values.

On this module we only study the first category. The second category is covered on the level 5 Business Statistics module and includes the ubiquitous **normal distribution**.

Before we can formally discuss probability distributions we need to be familiar with **random variables**.

8.1 Random Variable

Definition 31 (Discrete) Random Variable A random variable is a function whose domain is the sample space of an experiment and codomain is the real numbers. That is, a random variable converts an outcome of an experiment into a real number. It thus converts a value which may be non-numeric into a numeric value. For example,

- when tossing a coin we may use the random variable `numberOfHeads` that takes the outcome of the experiment – a head or a tail – and returns a 1 or a 0; and
- when tossing a coin twice we may use the random variable `totalNumberOfHeads` that takes the outcome of the experiment and returns the total number of heads.

-
1. Define a random variable for the following experiments.
 - (a) Selecting a student from the class.
 - (b) Rolling a dice.
 - (c) Rolling a dice twice.
 - (d) Introducing a new Masters programme in the Business School.
 2. Is a random variable *random* and/or *variable*?
-

8.2 Probability Distribution

A probability distribution simply assigns a probability to each value of a random variable. More formally (and correctly) we should say that a **discrete probability distribution** assigns a probability to each value of a **discrete random variable**. When we have a **continuous random variable** the variable can have infinitely many values each of which would have a probability of zero! In the continuous case instead of finding the probability of a single value, we find the probability of being within a range of values.

The function that assigns a probability to each random variable value is called a **probability mass function** which we typically, and unimaginatively, call f . A probability distribution is defined by associating the values of a random variable – typically called X – with their probability. Note that X is a function and so the values of X are actually the result of the application of X to an outcome of the experiment. The probabilities are generated by applying the probability mass function to the random variable value – $f(X)$.

Here are some example probability distributions presented in tabular form, with the values of the random variable in the first column alongside their probability in the second column.

Tossing A Coin Once

X	$f(X)$
0	0.5
1	0.5

Tossing A Coin Twice

X	$f(X)$
0	0.25
1	0.5
2	0.25

Rolling a Dice

X	$f(X)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Each probability distribution looks exactly like a **relative frequency distribution**. In common with a relative frequency distribution, a probability distribution is only **valid** if the sum of the probabilities is 1. In addition, we can generate more interesting probabilities directly from a probability distribution. For example,

- the probability of getting at most a 3 on a dice is the sum of the probabilities of getting 1, 2 or 3. This is because the event is the union of mutually exclusive events – getting a 1, getting a 2 and getting a 3 – and the addition rule states that you should then sum the individual probabilities;

- the probability of getting at least 2 on a dice is the same as getting anything except a one. Thus we can simply subtract the probability of getting a one from 1. That is, $1 - f(1)$. So instead of having to sum 5 probabilities we can simply subtract one probability from 1.

1. Write in tabular form the following probability distributions.
 - (a) Selecting a card from a pack of cards where the random variable associates 0 with a heart, 1 with a diamond, 2 with a club and 3 with with a spade.
 - (b) Rolling a dice twice where the random variable is `sumOfFaces`.
 - (c) Introducing a new Masters programme in the Business School. The random variable is `numberOfSuccesses`.
2. Using the probability distribution from question 1b, what is the probability of getting
 - (a) 2;
 - (b) 3;
 - (c) 2 or 3;
 - (d) at least 4.

In the following section we present a collection of named discrete probability distributions. In each case they have a standard form for their probability mass function and may be applied in various real world situations. We present each with their probability mass function, mean and variance.

8.3 Some Important Discrete Probability Distributions

Before we describe the various distributions, we present expressions that are used to generate the mean and variance for each distribution. We know from section 6 that the mean of a set of values can be calculated by summing the product of each data value and its relative frequency. Since a probability distribution is simply a special relative frequency distribution, its mean can be calculated in the same way. That is

$$\mu = x_1 * f(x_1) + x_2 * f(x_2) + \dots + x_n * f(x_n)$$

where x_i represents the random variable values.

This can be written using the summation notation \sum as

$$\mu = \sum_{i=1}^n x_i * f(x_i)$$

The variance, which we know is the mean of the squared differences of the data values from the mean, can also be calculated in the same way as for a relative frequency distribution. That is, using the summation notation

$$\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 * f(x_i)$$

We will use these expressions to generate the mean and variance for some of the following distributions, and for others we will simply present their values.

8.3.1 Discrete Uniform Distribution

A distribution where each value has an equal probability is a **discrete uniform distribution**. The random variable values will be a contiguous set of integers such as $\{0, 1, 2, 3\}$ or $\{4, 5, 6, 7, 8, 9\}$. If there are n random variable values then the probability of each value is $\frac{1}{n}$ (since the sum of the probabilities must equal 1). The probability mass function for a discrete uniform distribution with n values is therefore

$$f(x_i) = \frac{1}{n}, \forall x_i$$

The mean of a discrete uniform distribution with n values where the minimum value is a is

$$\begin{aligned} \mu &= \frac{1}{n} * a + \frac{1}{n} * (a + 1) + \dots + \frac{1}{n} * (a + (n - 1)) \\ &= \frac{1}{n} (a + (a + 1) + \dots + (a + (n - 1))) \\ &= \frac{1}{n} * \frac{n}{2} * (2a + (n - 1)) \\ &= \frac{2a + (n - 1)}{2} \end{aligned}$$

The variance of a discrete uniform distribution with n values is

$$\sigma^2 = \frac{n^2 - 1}{12}$$

Some example experiments that are modelled by a discrete uniform distribution are:

- tossing a coin once where the random variable is `numberOfHeads`. Here $a = 0$, $n = 2$ and the set of values is $\{0, 1\}$, and therefore, the mean is $\frac{2*0+1}{2} = 0.5$ and the variance is $\frac{2^2-1}{12} = 0.25$; and
- rolling a dice where the random variable is `faceNumber`. Here $a = 1$, $n = 6$ and the set of values is $\{1, 2, 3, 4, 5, 6\}$, and therefore, the mean is $\frac{2*1+5}{2} = 3.5$ and the variance is $\frac{6^2-1}{12} = \frac{35}{12}$.

Write down two more experiments that can be modelled by a discrete uniform distribution. In each case define the random variable and calculate its mean and variance.

8.3.2 Bernoulli Distribution

A Bernoulli distribution is a distribution whose random variable has two values – 0 and 1 (often referred to as **failure** and **success**) – and whose probability mass function is defined as

$$\begin{aligned}f(1) &= p \\f(0) &= 1 - p\end{aligned}$$

where p is the probability of success.

It is named after Jacob Bernoulli one of the many famous Bernoulli mathematicians. They include:

- his younger brother Johann Bernoulli who worked on calculus; and
- his nephew Daniel Bernoulli who worked on fluid mechanics.

The mean of a Bernoulli distribution is

$$\begin{aligned}\mu &= \sum_{i=1}^n x_i * f(x_i) \\&= 0 * (1 - p) + 1 * p \\&= p\end{aligned}$$

The variance of a Bernoulli distribution is

$$\begin{aligned}\sigma^2 &= \sum_{i=1}^n (x_i - \mu)^2 * f(x_i) \\&= (0 - p)^2 * (1 - p) + (1 - p)^2 * p \\&= p(1 - p)\end{aligned}$$

Some example experiments that are modelled by a Bernoulli distribution are:

- tossing a coin once. The random variable is `numberOfHeads`. Here $p = 0.5$ and thus the mean is 0.5 and variance is 0.25. Since we used this same example for the previous distribution, this experiment can be modelled by both a Bernoulli distribution and a discrete uniform distribution; and
- picking a student from the class. The random variable is `numberOfMales`. Using the data we have been using throughout this document $p = 0.7$, and thus the mean is 0.7 and the variance is 0.21.

Write down two more experiments that can be modelled by a Bernoulli distribution. In each case define the random variable and calculate its mean and variance.

8.3.3 Binomial Distribution

The binomial distribution is the distribution of the number of successes in n independent success or failure experiments. That is, it is the distribution of n independent Bernoulli experiments. For example, if you were to toss a coin several times (or several coins once) then you are doing repeated Bernoulli experiments. They are independent since the outcome of one does not impact on the outcome of any other. We typically call these **Bernoulli trials** rather than Bernoulli experiments.

Assuming that the probability of success on each Bernoulli trial is p , the probability mass function is

$$f(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Now $\binom{n}{x}$ represents the

- **number of ways of choosing x things from n ; or**
- **the number of combinations of x chosen from n ; or in this case**
- **the number of ways of getting x successes in n trials.**

It is known as a **binomial coefficient** and it is worthwhile dedicating a special section to its calculation.

Binomial Coefficients

We will explain the calculation of Binomial coefficients by using a simple example. Let's assume that there are 3 trials and we want to know how many ways we can get:

- 0 successes;
- 1 success;
- 2 successes; and
- 3 successes.

Let's make this even easier to understand by assuming that a trial is tossing a coin and a success is a head. Then we want to know in how many ways we can get:

- 0 heads;
- 1 head;
- 2 heads; and
- 3 heads.

- there is only 1 way to get 0 heads, if we get 3 tails. That is, TTT.
- There are 3 ways to get 1 head: HTT, THT, TTH.
- There are 3 ways to get 2 heads: HHT, HTH, THH.
- There is only one way to get 3 heads: HHH.

Note that there are 8 – or 2^3 – possibilities in total since for each toss we have two possibilities. The pattern for 0, 1, 2 and 3 successes is 1, 3, 3, 1, which is a pattern we have seen before in section 2.2.2 when we studied Pascal's Triangle!

Thus, if a set has n elements and an experiment involves n independent Bernoulli trials, then the number of subsets with 0, 1, 2 etc. elements is the same as the number of ways of getting 0, 1, 2 etc. successes in the n trials. But why is this? Well, if we view what happens to each element as a trial, where being in a subset is a success and not being in a subset is a failure, then the empty set represents 0 successes, each singleton set represents 1 success and so on.

Hence we can use Pascal's triangle to find $\binom{n}{x}$ in the same way as we used it to find the number of subsets with x elements of a set with n elements. In actual fact the values in each row of Pascal's Triangle are typically called **binomial coefficients** the name given to the values generated by $\binom{n}{x}$.

If you don't want to rely on Pascal's Triangle to calculate binomial coefficients there is an alternative way. Let's now use a similar example to the previous one except that we are tossing the coin 5 times (or tossing 5 coins once). In how many ways could we get 2 successes? That is, in how many ways can we choose 2 of the 5 tosses to have heads?

Well

- we have 5 choices for the first head – any of the tosses; and
- we have 4 choices for the second head – any of the remaining 4 tosses.

Now this suggests that there are 20 ($5 * 4$) ways to get 2 heads out of 5. However, these include:

- choosing the first toss and the second toss; and
- choosing the second toss and the first toss.

This may sound a bit strange – how can we choose the second toss for the first head and the first toss for the second head! Since tossing a coin 5 times is the same as tossing 5 coins simultaneously this choice is not strange – we are simply picking 2 out of a row of 5.

So clearly it doesn't matter which toss we choose for the first head – overall the same combination of tosses had heads. That is, **the order of the selections is not important**.

If it were important it would be called a **permutation**.

Thus each combination of selections occurs twice (**the number of ways that two things can be arranged**), and so, the number of ways that we can get 2 heads is not 20 but 10.

Therefore the number of ways of choosing 2 of the 5 tosses to have heads is

$$\frac{5 * 4}{2} = 10$$

This value of course applies to any situation where we want to select 2 things from 5 and the order that they are chosen does not matter.

We can now generalise the expression to one that applies when we want to select x things from n . We get

$$\frac{n * (n - 1) * \dots * (n + 1 - x)}{x * (x - 1) * \dots * 1}$$

where the numerator, $n * (n - 1) * \dots * (n + 1 - x)$, represents the product of the number of choices at each selection – as in $5 * 4$ – and the denominator, $x * (x - 1) * \dots * 1$, represents the number of ways of arranging x things.

Now assuming that $m! = m * (m - 1) * \dots * 1$ (we read $m!$ as **m factorial**) then the number of ways of selecting x things from n is

$$\frac{n!}{(n - x)! * x!}$$

-
1. What is the number of ways of arranging
 - (a) 3 things;
 - (b) 4 things;
 - (c) 5 things?
 2. What is the number of ways of selecting (combinations of) 3 things from
 - (a) 4 things;
 - (b) 5 things?
 3. What is the number of ways of selecting a set of National Lottery numbers?
-
-

Now we return to the probability mass function for the binomial distribution.

$$f(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Its value is:

- the product of the probabilities of x successes, p^x , multiplied by
- the product of the probabilities of $n - x$ failures, $(1 - p)^{n-x}$, multiplied by
- the number of different ways that this can be done – the binomial coefficient $\binom{n}{x}$.

Thus, when tossing a coin 3 times, the probability of

- 0 heads is

$$f(0) = \binom{3}{0} 0.5^0 (0.5)^3 = 1 * 1 * 0.125 = 0.125$$

- 1 head is

$$f(1) = \binom{3}{1} 0.5^1 (0.5)^2 = 3 * 0.5 * 0.25 = 0.375$$

- 2 heads is

$$f(2) = \binom{3}{2} 0.5^2 (0.5)^1 = 3 * 0.25 * 0.5 = 0.375$$

- 3 heads is

$$f(3) = \binom{3}{3} 0.5^3 (0.5)^0 = 1 * 0.125 * 1 = 0.125$$

The mean of a binomial distribution with n trials and a probability of success of p is np . The variance is $np(1 - p)$. Thus the mean for the above example is 1.5 and the variance is 0.75.

Some example experiments that are modelled by a Binomial distribution are:

- selecting 5 cards from a pack of cards where the card is replaced after each selection. The random variable is `numberOfRedCards`. Since $n = 5$ and $p = 0.5$, the mean is 2.5 and the variance is 1.25.
- checking 100 products for faults where the probability of a faulty product 0.05. The random variable is `numberOfFaultyProducts`. Since $n = 100$ and $p = 0.05$, the mean is 5 and the variance is 4.75.
- checking 20 customers for satisfaction with a Website where the probability of being satisfied is 0.8. The random variable is `numberOfSatisfiedCustomers`. Since $n = 20$ and $p = 0.8$, the mean is 16 and the variance is 3.2.

1. For the random variable `numberOfRedCards` what is the probability of

- (a) 0 red cards;
- (b) 1 red card;
- (c) at least 1 red card.

2. For the random variable `numberOfSatisfiedCustomers` what is the probability of

- (a) 2 satisfied customers;
- (b) 0 or 20 satisfied customers.

Now many experiments involve counting the number of successes where the probability of success is not constant. For example, if one wants to count the number of red cards in a selection of cards, where the cards are not replaced after each selection, then the probability of selecting a red card (a success) changes at each selection.

This type of experiment is modelled by a **hypergeometric distribution**.

8.3.4 Hypergeometric Distribution

We will illustrate the use of the hypergeometric distribution with an example. Let's assume that we have 10 students, 6 of whom are female and 4 are male. If we randomly select 3 students, what is the probability of selecting

- 0 females;
- 1 female;
- 2 females; and
- 3 females

Since we are counting the number of females, we are equating female with success. It is clear however that at each selection the probability of success is different. The probability of

- 0 females is the probability of selecting a male on each selection. This is

$$\frac{4}{10} * \frac{3}{9} * \frac{2}{8} = 0.03$$

- 1 female is the probability of selecting 1 female (F) and 2 males (M). Now there are three ways this can be done – FMM, MFM, and MMF – as highlighted in the following calculation

$$\left(\frac{6}{10} * \frac{4}{9} * \frac{3}{8}\right) + \left(\frac{4}{10} * \frac{6}{9} * \frac{3}{8}\right) + \left(\frac{4}{10} * \frac{3}{9} * \frac{6}{8}\right)$$

Note how each product of fractions contains the same numbers in the numerators and the denominators. Thus we get

$$\left(\frac{6}{10} * \frac{4}{9} * \frac{3}{8}\right) + \left(\frac{4}{10} * \frac{6}{9} * \frac{3}{8}\right) + \left(\frac{4}{10} * \frac{3}{9} * \frac{6}{8}\right) = 3 * \left(\frac{6}{10} * \frac{4}{9} * \frac{3}{8}\right) = 0.3$$

- 2 females is the probability of selecting 2 females and 1 male. There are also 3 ways to do this – FFM, FMF and MFF – and once again we can multiply a product of three fractions by 3. That is,

$$3 * \left(\frac{6}{10} * \frac{5}{9} * \frac{4}{8}\right) = 0.5$$

and

- 3 females is

$$\frac{6}{10} * \frac{5}{9} * \frac{4}{8} = 0.17$$

Note that the products of the fractions are multiplied by 1, 3, 3 and 1 respectively – the binomial coefficients again!

In summary, if we are selecting n things from N things **without replacement**, so that the probability of success changes with each selection, then we use the hypergeometric distribution. Now if there are m successes in the N things, and we want to calculate the probability of x successes, we are calculating the proportion of all possible selections of n from N that include x successes, selected from the m that are available. That is, we need to calculate

- the number of ways of selecting x successes from the m successes, $\binom{m}{x}$;
- the number of ways of selecting $n - x$ failures from the $N - m$ failures, $\binom{N-m}{n-x}$;
- the number of ways of selecting n things from the N things, $\binom{N}{n}$.

The probability is the ratio of the number of selections of n things from N things that include x successes, to the total number of selections of n things from N things. That is,

$$\frac{\binom{m}{x} * \binom{N-m}{n-x}}{\binom{N}{n}}$$

This is the probability mass function for the hypergeometric distribution. Its mean is $\frac{nm}{N}$ and its variance is $\frac{nm(N-n)(N-m)}{N^2(N-1)}$.

Returning to the example used at the beginning of this section we have $N = 10$, $m = 6$, and $n = 3$. Now the probability of

- 0 females can be calculated by letting $x = 0$,

$$\frac{\binom{6}{0} * \binom{4}{3}}{\binom{10}{3}} = \frac{1 * 4}{120} = 0.03$$

- 1 female can be calculated by letting $x = 1$,

$$\frac{\binom{6}{1} * \binom{4}{2}}{\binom{10}{3}} = \frac{6 * 6}{120} = 0.3$$

- 2 females can be calculated by letting $x = 2$,

$$\frac{\binom{6}{2} * \binom{4}{1}}{\binom{10}{3}} = \frac{15 * 4}{120} = 0.5$$

and

- 3 females can be calculated by letting $x = 3$,

$$\frac{\binom{6}{3} * \binom{4}{0}}{\binom{10}{3}} = \frac{20 * 1}{120} = 0.17$$

Since $n = 3$, $m = 6$ and $N = 10$, the mean is 1.8 and the variance is 0.15.

1. If we are selecting 3 cards from a pack of cards without replacement what is the probability of
 - (a) 0 red cards;
 - (b) 1 red card;
 - (c) at least 1 red card.
2. What is the mean number of red cards selected?
3. In a situation where one is selecting without replacement, when would it be appropriate to use the binomial distribution?