"Essential Mathematics & Statistics for Science" Second Edition by Dr G Currell & Dr A A Dowman (John Wiley & Sons)

Answers to 'Q' Questions

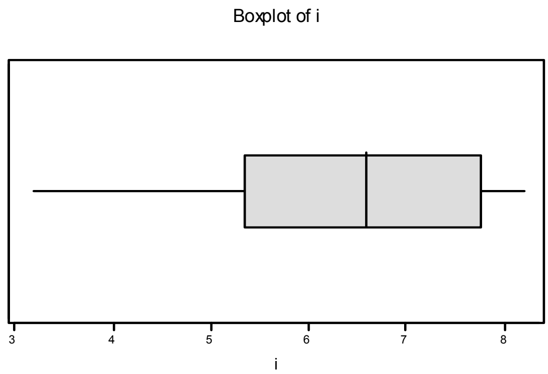
7 Statistics & Information

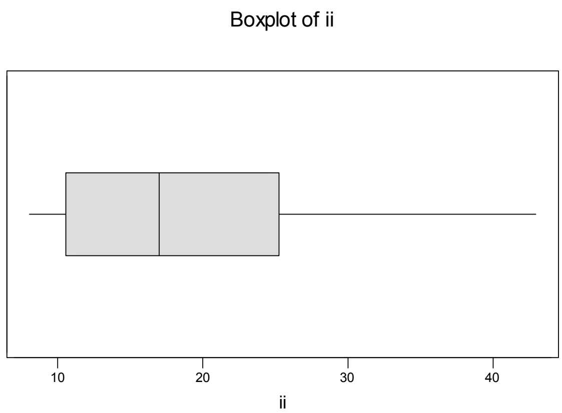
To navigate, use Bookmarks in the PDF file.

Q7.1

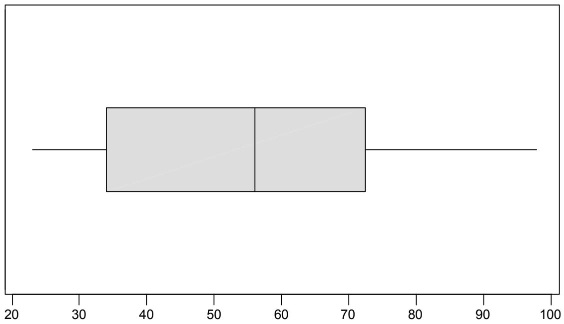
		Median	Lower Quartile	Upper Quartile	IQR Range
i)	5.0, 8.2, 7.9, 6.6, 7.6, 5.7, 3.2, 7.5, 5.9	6.6	5.35	7.75	2.4
ii)	11, 8,13, 9, 21, 24,12, 22, 29, 43	17	10.5	25.25	14.75
iii)	45, 67, 23, 78, 67, 56, 98, 23, 49	56	34	72.5	38.5

Q7.2





Boxplot of iii



See MINITAB file

Q7.3

The box and whisker plots of Figure 7.3 represent ALL the data values in each set including the maximum and minimum. The confidence intervals in Figure 7.4 show the statistics derived from each data set so that from one extreme to the other represents the 95% probability of the mean value lying between them.

Q 7.4

For each part of the question, you should imagine **repeating the process** by which the data set is defined - if the data values stay the same, then you have a population; if not, you have a sample.

i) **Population**: The four numbers in this set are specified. The process of identifying the numbers always gives the same set of values.

ii) **Sample**: If the experimental measurements are repeated a different set of four data values will be produced (although by coincidence some values could be the same).

iii) **Sample**: Spinning the roulette wheel 16 more times will give a new set of 16 possible values (although by coincidence some values could be the same).

iv) **Population**: The numbers on the face of the roulette wheel will be the same, each time you look.

v) **Population**: The scores are a matter of documented record and will not change.

vi) **Sample**: Repeating the survey should give a different random set of children who may have different levels of pocket money..

Q7.5

$$\sum_1^5 (x_i) = 4.9 + 4.7 + 5.1 + 4.9 + 4.4 = 24$$

Q7.6

i) $\bar{x} = \frac{\sum_1^n (x_i)}{n}$

Sum of all data values is given by:

$\sum_1^5 (x_i) = 4.9 + 4.7 + 5.1 + 4.9 + 4.4 = 24$

The number of data values:

n = 5

Hence mean value: $\bar{x}$ = 24/5 = 4.8

ii) $\bar{x} = \frac{\sum_1^n (x_i)}{n}$

Sum of all data values is given by:

$\sum_1^4 (x_i) = 0.25 + 0.27 + 0.19 + 0.22 = 0.93$

The number of data values:

n = 4

Hence mean value: $\bar{x}$= 0.93/4 = 0.2325

Q7.7

For the following set of x values:

$x_1 = 4.9$, $x_2 = 4.7$, $x_3 = 5.1$, $x_4 = 4.9$, $x_5 = 4.4$

i) Mean value, $\bar{x}$ = 4.8 (see previous question)

ii) Each deviation, $d_i = x_i - \bar{x}$

$d_1 = x_1 - \bar{x} = 4.9 - 4.8 = 0.1$

Similarly

$d_2 = -0.1$, $d_3 = 0.3$, $d_4 = 0.1$, $d_5 = -0.4$

iii) Sum of all deviations,

$\sum_1^n (d_i) = d_1 + d_2 + d_3 + d_4 + d_5 = 0.1 - 0.1 + 0.3 + 0.1 - 0.4 = 0$

iv) Yes - the sum of deviations is always zero.

This is because of the way in which the mean value is calculated.

v) Sum of all (deviations)^2,

$\sum_1^n (d_i^2) = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$

= 0.01 + 0.01 + 0.09 + 0.01 + 0.16 = 0.28

The sum of all (deviations)^2 will not be zero unless all the values are the same.

vi) $s^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n-1} = \frac{\sum (d_i)^2}{n-1} = \frac{0.28}{(5-1)} = 0.07$

vii) $s = \sqrt{s^2} = \sqrt{0.07} = 0.0265$

Q7.8

		Mean	Sample Variance	Sample Standard Deviation
i)	8, 6, 7, 4, 7	6.4	2.3	1.517
ii)	68, 66, 67, 64, 67	66.4	2.3	1.517
iii)	418, 416, 417, 414, 417	416.4	2.3	1.517

iv) All three data sets give the same variance and standard deviation, i.e. they all have the same *spread (or dispersion)* of data values.
However they have different mean values - their *location* on the number line is different.
Note that, when calculating ***variance*** **or** ***standard deviation*****, it is possible to add, or subtract, the same number from every data value in the set - you will still get the same answer.**

Q7.9

i)

Stem & Leaf Diagram	Number in each range
20 | 5 7	2
30 | 2 3 9	3
40 | 0 2 2 4 6 6 7 8 9 9	10
50 | 1 5 5 7 8	5
60 | 1 2 6 7	4

ii) The number of data values within each specified range is shown in answer (i)

Q7.10

i) See table below. The total of all frequencies must equal the number of graduates surveyed = 24.

ii) The relative frequency for each class, *g*, is the class frequency, *f*(*g*), divided by the total of all frequencies. See table below for results rounded to 3 significant figures.

iii) Total = 1.001.

iv) We would expect the total of all relative frequencies to be exactly 1.000. The reason for the slight difference is that the slight errors in rounding the different numbers added up to give a total error of 0.001

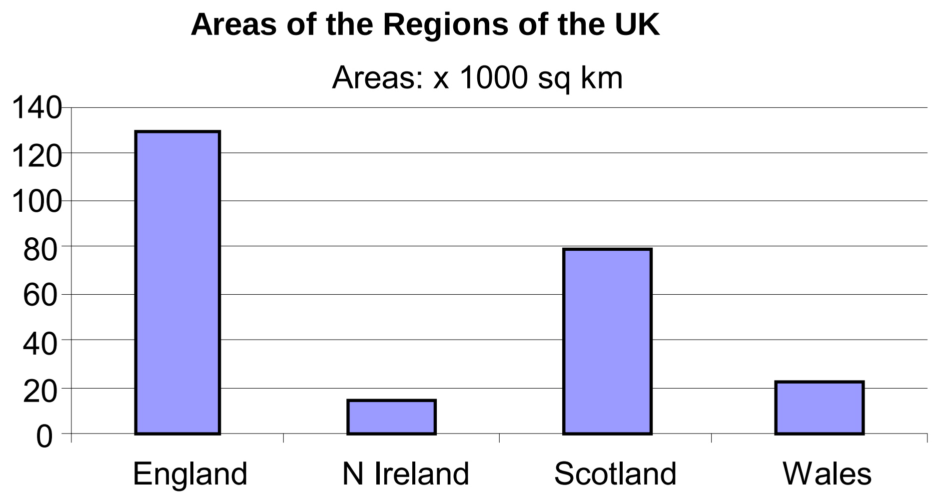
Salary range	Number (frequency)	Relative frequency
g	*f*(*g*)	
$12.0 \leq S < 16.0$	4	4/24 = 0.167
$16.0 \leq S < 20.0$	10	10/24 = 0.417
$20.0 \leq S < 22.0$	7	7/24 = 0.292
$24.0 \leq S < 28.0$	3	3/24 = 0.125
Totals:	24	1.001

Q7.11

Calculations of Fractions and Angles gives:

	Area ($\times$1000 km^2)	Fraction	Angle (degrees)
England	130	0.533	192
Northern Ireland	14	0.057	21
Scotland (and islands)	79	0.324	117
Wales	21	0.086	31
Total:	244	1.000	360

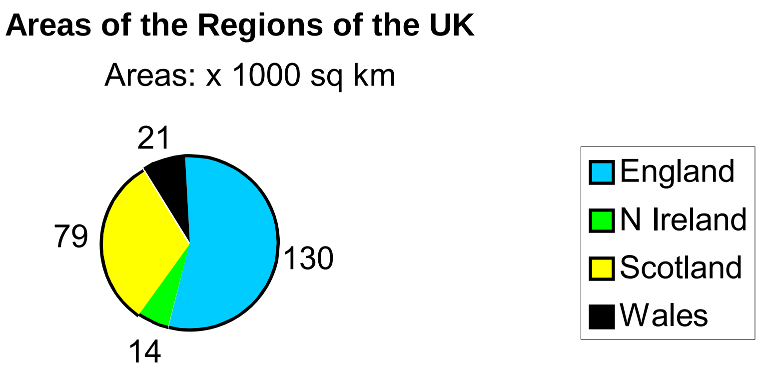
i)



ii) See above table

iii) See above table

iv)



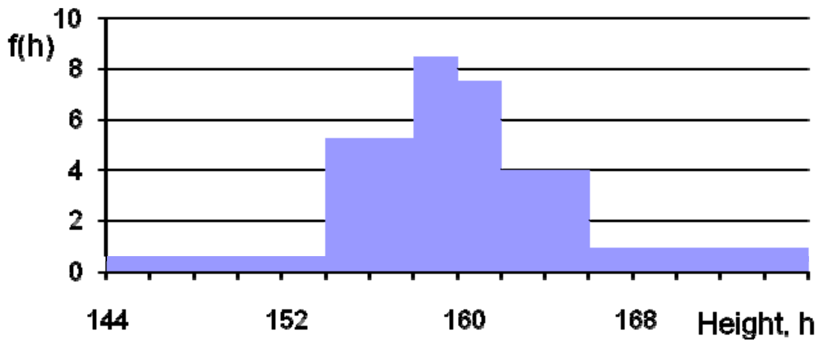
Q7.12

i)

Class Range h (cm)	Class Frequency $f_g = f(g)$	Class Width $\Delta x_g = \Delta x(g)$	Class Density $fd_g = fd(g)$
$144 < h \leq 154$	6	10	= 6 / 10 = 0.6
$154 < h \leq 158$	21	4	= 21 / 4 = 5.25
$158 < h \leq 160$	17	2	= 17 / 2 = 8.5
$160 < h \leq 162$	15	2	= 15 / 2 = 7.5
$162 < h \leq 166$	16	4	= 16 / 4 = 4.0
$166 < h \leq 176$	9	10	= 9 / 10 = 0.9

ii) Yes. The classes are continuous along a quantitative axis - the data can be drawn as a histogram.

iii)



Histogram of Frequency Density plotted against Height

The area of each 'bar' will give the number of children within the height range given along the horizontal axis.
For example the number of children with heights between 144 and 154 will be equal to $0.6 \times (154 - 144) = 0.6 \times 10 = 6$

Q7.13

i) See Table below

ii) See Table below

iii) See Table below
The Class Sum gives an estimate of the total salaries of all graduates in that class, e.g. there are 4 graduates in the class which has a mean class value of '14', hence a good estimate of the total = $4 \times 14 = 56$

iv) See Table below
Our estimate of the total salary of all graduates is £468,000

v) See Table below
Based on the total from (iv) the mean salary is £19,500

Salary Range	Frequency	Mean Class Value	Class Sum
Class, *g*	f_g	$\overline{S}_g$	$f_g \times \overline{S}_g$
$12.0 \leq S < 16.0$	4	14.0	56.0
$16.0 \leq S < 20.0$	10	18.0	180.0
$20.0 \leq S < 24.0$	7	22.0	154.0
$24.0 \leq S < 28.0$	3	26.0	78.0
$n = \sum_g (f_g) =$	24	$Sum = \sum_g (f_g \times \overline{S}_g) =$	468.0
	$Mean\ Value,\ \overline{S} = \dfrac{\sum_g (f_g \times \overline{S}_g)}{n} =$		19.5

vi) Adding all the salaries **individually** gives a true total of £461,500
This is close to the estimate in (iv), but differs slightly because we made the approximation in (ii) that all the graduates in any particular class had a salary equal to the mean class value.

vii) The true mean = £19,229, based on the true total in (vi) not the estimate in (iv)

viii) The method used in steps (i) to (v) must be used instead of the more accurate method used in (vi) and (vii) when the data is already grouped into classes and the individual values are no longer available, e.g. in a survey when respondents are only asked to give their salary within given ranges.

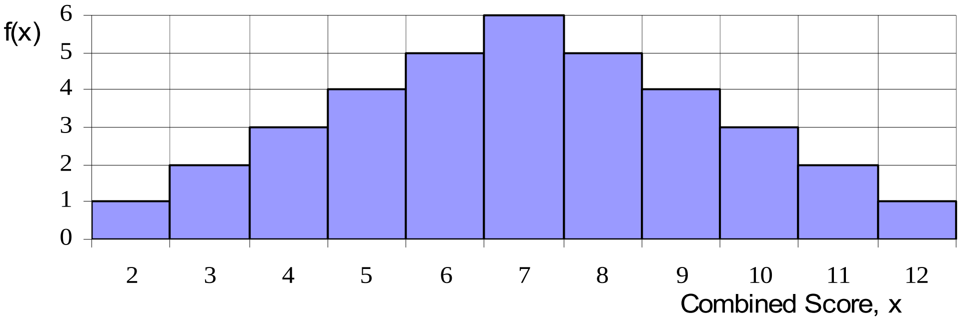
Q7.14

i) The number of ways are worked out in the Table below:

Combined Score, *x*	Different Ways	Number, *f*(*x*)
2	1+1	1
3	1+2, 2+1	2
4	1+3, 2+2, 3+1	3
5	1+4, 2+3, 3+2, 4+1	4
6	1+5, 2+4, 3+3, 4+2, 5+1	5
7	1+6, 2+5, 3+4, 4+3, 5+2, 6+1	6
8	2+6, 3+5, 4+4, 5+3, 6+2	5
9	3+6, 4+5, 5+4, 6+3	4
10	4+6, 5+5, 6+4	3
11	5+6, 6+5	2
12	6+6	1

Total number of ways = 36

ii)



iii) Total 'area' = 36

iv) The total 'area' equals the total number of different results that could be obtained when rolling the two dice.
Each die gives 6 possible values, hence the total combined number of possibilities = 6 × 6 = 36, which equals the area on your histogram.

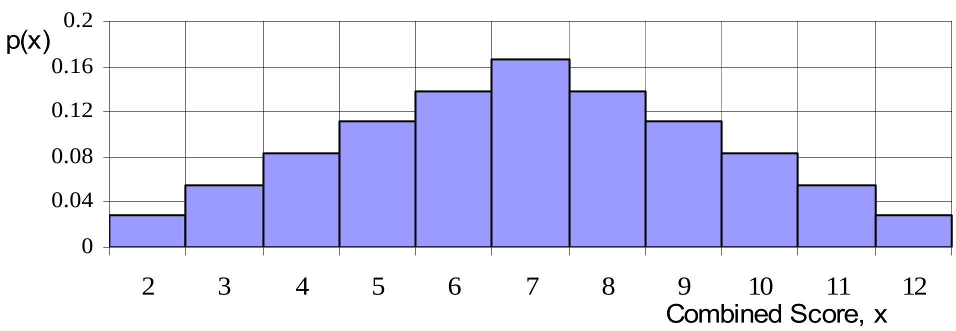
Q7.15

i) The probabilities for the different scores are calculated as follows:

Score, *x*	Frequency, *f*(*x*)	Probability, *p*(*x*)	
2	1	= 1 / 36 =	0.0278
3	2	= 2 / 36 =	0.0556
4	3	= 3 / 36 =	0.0833
5	4	= 4 / 36 =	0.1111
6	5	= 5 / 36 =	0.1389
7	6	= 6 / 36 =	0.1667
8	5	= 5 / 36 =	0.1389
9	4	= 4 / 36 =	0.1111
10	3	= 3 / 36 =	0.0833
11	2	= 2 / 36 =	0.0556
12	1	= 1 / 36 =	0.0278
Total Probability =			1.0001

(The total area actually equals 1.0000, but the rounding errors in the individual probabilities cause a small error in the final decimal place of the total)

ii)



iii) Total 'area' = 1.00

iv) The total 'area' equals the probability of getting one of all the possible different results that could be obtained when rolling the two dice. Since it is certain that you will get one of the possible results, the probability must be equal to '1'.

Q7.16

The area under the graph between 30% and 40% is approximately equal to a triangle with a height, 0.01, and a width, 10.

The area (= half × base × height) of this triangle would be 0.5 × 0.01 × 10 = 0.05.

Hence there is a 0.05 (= 5%) probability (approximately) that a random student will get a mark between 30% and 40%.

Q7.17

i) For each new value of *x*, add up the values of *p*(*x*) for every value of *x* up to an including the new value:
(The values for *p*(*x*) can be found in the answer to Q7.15)

Score	Probability	Cumulative Probability
x	***p*(*x*)**	***cp*(*x*)**
2	0.0278	0.0278
3	0.0556	0.0833
4	0.0833	0.1667
5	0.1111	0.2778
6	0.1389	0.4167
7	0.1667	0.5833
8	0.1389	0.7222
9	0.1111	0.8333
10	0.0833	0.9167
11	0.0556	0.9722
12	0.0278	1.0000

ii) The probability of scoring 6 or less can be read directly from the table or the graph.
p(6 or less) = 0.4167

iii) The probability of scoring at least 10 is equal to 1 minus the probability of scoring 9 or less:
p(10 or more) = 1.0000 - *p*(9 or less) = 1.0000 - 0.8333 = 0.1667

Q7.18

Grade	Frequency	Relative	Probability	Number	Integer

		Frequency			**Numbers**
First Class:	18	0.1286	0.1286	3.8571	4
Upper Second:	38	0.2714	0.2714	8.1429	8
Lower Second:	50	0.3571	0.3571	10.7143	11
Third Class:	34	0.2429	0.2429	7.2857	7
Totals =	140	1		30	30

i)	Total = Sum of all Frequency Values = 140 Relative Frequency of Class = Frequency of Class / Total e.g. Relative Frequency for First Class = 18 / 40 = 0.1286
ii)	*Future* Probability for Class = Relative Frequency of Class **in** *past*. *Future* Probability for First Class = 0.1286
iii)	Predicting the future probability on past relative frequency requires that the conditions of the system do not change. In this case we make the assumption that the capabilities of future students is similar to those in the past and that the effectiveness of the teaching remains broadly similar.
iv)	Future numbers (frequency) in Class = Total Number × Probability in Class e.g. Future Frequency for First Class = 30 × 0.1286 = 3.8571 (similarly for the other classes of degree) As it is not possible to have 0.8571 students, it would be appropriate to round the predicted numbers to integer values.

Q7.19

i)	The king of clubs is a **single** card selected from a possible 52. Probability, *p*(king of clubs) = 1/52
ii)	There are 4 possible kings: Probability, *p*(any king of clubs) = *p*(king of clubs) + *p*(king of diamonds) + *p*(king of hearts) + *p*(king of spades) = 1/52 + 1/52 +1/52 +1/52 = 4/52 = 1/13
iii)	There are 12 possible picture cards: Probability, *p*(any picture card) = 12 × 1/52 = 3/13
iv)	There are 4 possible aces: Probability, *p*(any ace) = *p*(ace of clubs) + *p*(ace of diamonds) + *p*(ace of hearts) + *p*(ace of spades) = 1/52 + 1/52 +1/52 +1/52 = 4/52 = 1/13
v)	There are 13 possible cards of the club suit: Probability, *p*(any club) = 13 × 1/52 = 1/4

Q7.20

P(tall) = 0.75
P(short) = 0.25
P(round) = 0.75
P(wrinkled) = 0.25

i) P(tall and round) = P(tall) × P(round) = 0.75 × 0.75 = 0.5625

ii) P(short and round) = P(short) × P(round) = 0.25 × 0.75 = 0.1875

iii) P(tall and wrinkled) = P(tall) × P(wrinkled) = 0.75 × 0.25 = 0.1875

iv) P(short and wrinkled) = P(short) × P(wrinkled) = 0.25 × 0.25 = 0.0625

Note that the addition of all probabilities in (i) to (iv) gives a total of '1' - there are only the four possibilities when selecting one of the plants.

Q7.21

P(A) = 1/8, P(B) = 1/6, P(C) = 1/3, P(D) = 9/24.

i) P(A or B) = P(A) + P(B) = 1/8 + 1/6 = (3 + 4)/24 = 7/24

ii) P(A *or* B or C) = P(A) + P(B) + P(C) = 1/8 + 1/6 + 1/3
= (3 + 4 + 8)/24 = 15/24

iii) P(A or B or C or D) = P(A) + P(B) + P(C) + P(D) = 1/8 + 1/6 + 1/3 + 9/24
= (3 + 4 + 8 + 9)/24 = 24/24 = 1
The probability of '1' means that one of A or B or C or D must occur.

Q7.22

i) P(king of clubs) = P(king of diamonds) = 1/52
P(king of clubs or king of diamonds) = 1/52 + 1/52 = 2/52 = 1/26

ii) P(king of hearts or king of clubs or king of diamonds or king of spades) = 1/52 + 1/52 + 1/52 + 1/52
= 4/52 = 1/13

iii) P(any picture card) = 12/52
P(any ace) = 4/52
Hence:
P(any picture card or ace) = 12/52 + 4/52 = 16/52 = 4/13

Q7.23

i) P(king) = 1/52 +1/52 +1/52 +1/52 = 1/13

ii) P(NOT king) = 1 - P(king) = 1 - 1/13 = 12/13

iii) P(king or NOT king) = P(king) + P(NOT king) = 1/13 + 12/13 = 13/13 =1
Every card that is picked will either be a king or NOT a king.

Q7.24

Probability of any particular score, x, on one die, $P(x) = 1/6$

i) A total of '2' on two dice can only be obtained in one way: 1 + 1
$P(\text{total } 2) = P(1) \times P(1) = 1/6 \times 1/6 = 1/36 = 0.027778$

ii) A total of '3' on two dice can be obtained in two ways: 1 + 2 and 2 + 1
$P(\text{total } 3) = P(1)\times P(2) + P(2)\times P(1) = 1/6\times 1/6 + 1/6\times 1/6 = 2\times(1/36) = 0.055556$

iii) A total of '7' on two dice can be obtained in six ways:
1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1
$P(\text{total } 7) = 6 \times (1/36) = 0.16667$

iv) A total of '12' on two dice can only be obtained in one way: 6 + 6
$P(\text{total } 12) = P(6) \times P(6) = 1/6 \times 1/6 = 1/36 = 0.027778$

Q7.25

Probability that any **one** seed will germinate = 0.9

i) Probability that any **one** seed **out of one** will germinate:
$P(1) = 0.9$

ii) Probability that any **one** seed **out of one** will NOT germinate:
$P(0) = 1 - P(1) = 0.1$

iii) Probability that **four** seeds **out of four** will germinate:
$P(1,1,1,1) = 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.6561$

iv) Probability that **no** seeds **out of four** will germinate:
$P(0,0,0,0) = 0.1 \times 0.1 \times 0.1 \times 0.1 = 0.0001$

v) The germination of **one** seed **out of four** can occur in 4 different ways:
$P(1 \text{ out of } 4) = P(1,0,0,0) + P(0,1,0,0) + P(0,0,1,0) + P(0,0,0,1)$
The probability of each is the same:
$P(1,0,0,0)=P(0,1,0,0)=P(0,0,1,0)=P(0,0,0,1) = 0.9 \times 0.1 \times 0.1 \times 0.1 = 0.0009$
Thus:
$P(1 \text{ out of } 4) = 4 \times 0.0009 = 0.0036$

vi) The germination of **three** seeds **out of four** can occur in 4 different ways:
$P(1 \text{ out of } 4) = P(1,1,1,0) + P(1,1,0,1) + P(1,0,1,1) + P(0,1,1,1)$
The probability of each is the same:
$P(1,1,1,0)=P(1,1,0,1)=P(1,0,1,1)=P(0,1,1,1) = 0.9 \times 0.9 \times 0.9 \times 0.1 = 0.0729$
Thus:
$P(1 \text{ out of } 4) = 4 \times 0.0729 = 0.2916$

Q7.26

A man who tests negative could either have a true negative or a false negative result.
For a man, selected at random, to record a true negative result requires that the man does not have the disease AND he must then test negative.

Probability that a man, selected at random, does not have the disease, P($\overline{\mathrm{D}}$) = 0.95

- Probability that a man, who does not have the disease, will test negative, P($\overline{\mathrm{P}}\,|\,\overline{\mathrm{D}}$) = 0.8

Hence, the probability that a man, selected at random, will record a true negative result

P(True negative) = P($\overline{\mathrm{P}}\,|\,\overline{\mathrm{D}}$) $\times$ P($\overline{\mathrm{D}}$) = 0.95 $\times$ 0.8 = 0.76

For a man, selected at random, to record a false negative result requires that the man does have the disease AND he must then test negative.

- Probability that a man, selected at random, has the disease, P(D) = 0.05
- Probability that a man, with the disease, will test negative, P($\overline{\mathrm{P}}$|D) = 0.1

Hence, the probability that a man, selected at random, will record a false negative result

P(False negative) = P($\overline{\mathrm{P}}$|D) $\times$ P(D) = 0.05 $\times$ 0.1 = 0.005

Hence, for a man, selected at random, who tests negative for the disease, the probability that he does actually have the disease is given by:

$$P(\textit{Has the disease}) = \frac{P(\textit{False negative})}{P(\textit{True negative}) + P(\textit{False negative})} = \frac{0.005}{0.765} = 0.0065$$

Q7.27

i) Probability of selecting a black sweet from 3 black and 4 red is:
P(B|3B4R) = 3/7
Then there will be 2 black and 4 red left, and probability of selecting a black sweet from 2 black and 4 red is
P(B|2B4R) = 2/6
The combined probability is then
P(B,B) = *P*(B|3B4R) $\times$ *P*(B|2B4R) = 3/7 $\times$ 2/6 = 1/7 = 0.1429

ii) Probability of selecting a black sweet from 3 black and 4 red is:
P(B|3B4R) = 3/7
Then there will be 2 black and 4 red left, and probability of selecting a red sweet from 2 black and 4 red is
P(R|2B4R) = 4/6
The combined probability is then
P(B,R) = *P*(B|3B4R) $\times$ *P*(R|2B4R) = 3/7 $\times$ 4/6 = 2/7 = 0.2857

iii) Probability of selecting a black sweet from 3 black and 4 red is:
P(R|3B4R) = 4/7
Then there will be 3 black and 3 red left, and probability of selecting a red sweet from 3 black and 3 red is
P(R|3B3R) = 3/6 = 1/2
The combined probability is then
P(R,R) = P(R|3B4R) × P(R|3B3R) = 4/7 × 1/2 = 2/7 = 0.2857

iv) Probability of selecting a black sweet from 3 black and 4 red is:
P(R|3B4R) = 4/7
Then there will be 3 black and 3 red left, and probability of selecting a black sweet from 3 black and 3 red is
P(B|3B3R) = 3/6 = 1/2
The combined probability is then
P(R,B) = P(R|3B4R) × P(B|3B3R) = 4/7 × 1/2 = 2/7 = 0.2857

v) Probability of any one of the above four combinations
= P(B,B) + P(B,R) + P(R,R) + P(R,B) = 1/7 + 2/7 + 2/7 + 2/7 = 7/7 = 1

The above four combinations are the only ways of selecting two sweets one after the other. It is certain that one of the four combinations will occur - hence the combined probability of '1'.

Q7.28

We will first calculate the probability that no-one in 5 people has their birthday in the same month as anyone else.

Pick the first person.
Then the probability that the *next* person (person 2) has a birthday in a different month (i.e. has a birthday in one of the 11 months different from person 1):
P(2 different) = 11/12
Then pick person 3, who must have a birthday in one of the 10 months different from persons 1 and 2:
P(3 different) = 10/12
Similarly for person 4 (there are only 9 months left for his/her birthday):
P(4 different) = 9/12
And then for the last person, 5, only 8 months are left:
P(5 different) = 8/12

The combined probability that are ALL five are different is given by:
P(All five different)
= P(2 different) × P(3 different) × P(4 different) × P(5 different)
= 11/12 × 10/12 × 9/12 × 8/12 = 0.382

Hence the probability that AT LEAST two people have a birthday in the same month is given by
P(At least two the same) = 1 - P(All five different) = 1 - 0.382 = 0.618

Q7.29

i) $5! = 5\times4\times3\times2\times1 = 120$

ii) $(4 - 3)! = 1! = 1$

iii) $(3 - 3)! = 0! = 1$

iv) $3! - 3! = 0$

v) $$\frac{7!}{5!} = \frac{7\times6\times5\times4\times3\times2\times1}{5\times4\times3\times2\times1} = 7\times6 = 42$$

vi) $$\frac{101!}{99!} = \frac{101\times100\times99!}{99!} = 101\times100 = 10{,}100$$

Q7.30

Number of ways of arranging 4 different objects = 4! = 24

There are 4 choices for the first place, then, for each of these 4 choices, there are 6 ways of rearranging the next 3:

ATGC, ATCG, ACTG, ACGT, AGCT, AGTC
TAGC, TACG, TCAG, TCGA, TGAC, TGCA
CATG, CAGT, CGAT, CGTA, CTAG, CTGA
GCTA, GCAT, GATC, GACT, GTAC, GTCA

Q7.31

There are 8 choices for the first trophy place, then 7 choices for the next place, 6 choices for the third, and 5 choices for the fourth trophy place.

Number of different ways = $8\times7\times6\times5 = 1680$

Using the Permutations formula, the number of ways of ordering $r = 4$ items from a total, $n = 8$, is given by:

${}_nP_r = {}_8P_4 = 8!\,/\,4! = 1680$

Q7.32

In selecting 11 players, the order of selection is not important.
The selection of 11 players from the squad of 20 is then given by the Combination function:

Number of ways = ${}_nC_r$ = ${}_{20}C_{11}$ = 167960