Answers to Sheet 1

Types of Statistical Data and Display of Data

1. Discrete means it's a number usually obtained from counting. Continuous means any value with lots of decimals places is possible if we could measure it.
   a. Continuous. Times are usually continuous. With accurate equipment we could measure to fractions of a second.
   b. Discrete.
   c. Discrete. Prices are discrete.
   d. Categorical.
   e. Discrete.
   f. Continuous.
   g. Categorical. Grades are A, B, C etc.
   h. Continuous.
2.
   a. The data is discrete since we are counting how many times something has happened.
   b.

Eggs	Frequency	Relative Frequency	Cumulative Frequency
25-29	2	0.07	2
30-34	5	0.17	7
35-39	4	0.13	11
40-44	6	0.2	17
45-49	3	0.1	20
50-54	6	0.2	26
55-59	3	0.1	29
60-64	1	0.03	30

3
   a. The data is continuous because we are measuring something. With perfect measuring equipment any degree of accuracy is possible i.e. lots of decimal places.

Measured length	Boundaries	Frequency	Relative Frequency	Cumulative Frequency
168-169	167.5-169.5	3	0.1	3
170-171	169.5-171.5	3	0.1	6
172-173	171.5-173.5	3	0.1	9
174-175	173.5-175.5	2	0.07	11
176-177	175.5-177.5	2	0.07	13
178-179	177.5-179.5	5	0.17	18
180-181	179.5-181.5	4	0.13	22
182-183	181.5-183.5	2	0.07	24
184-185	183.5-185.5	2	0.07	26
186-187	185.5-187.5	4	0.13	30

b.

Measured length	Boundaries	Frequency	Relative Frequency	Cumulative Frequency
168-171	167.5-171.5	6	0.2	6
172-175	171.5-175.5	5	0.17	11
176-179	175.5-179.5	7	0.23	18
180-183	179.5-183.5	6	0.2	24
184-187	183.5-187.5	6	0.2	30

c. Altering the way we round alters the class boundaries. See the next question for an explanation of this. We get the following table.

Measured length	Boundaries	Frequency	Relative Frequency	Cumulative Frequency
168-171	168-172	6	0.2	6
172-175	172-176	5	0.17	11
176-179	176-180	7	0.23	18
180-183	180-184	6	0.2	24
184-187	184-188	6	0.2	30

4.

a.

Measured length	Boundaries	Frequency	Relative Frequency	Cumulative Frequency
50-51	49.5-51.5	7	0.14	7
52-53	51.5-53.5	4	0.08	11
54-55	53.5-55.5	12	0.24	23
56-57	55.5-57.5	5	0.1	28
58-59	57.5-59.5	1	0.02	29
60-64	59.5-64.5	2	0.04	31
65-69	64.5-69.5	2	0.04	33
70-79	69.5-79.5	5	0.1	38
80-99	79.5-99.5	5	0.1	43
100 -149	99.5-149.5	3	0.06	46
150 - 199	149.5-199.5	4	0.08	50

b. If the data is recorded by rounding down to the nearest mm, what values will be recorded as 184? The smallest value to be recorded as 184 will be 184 because anything slightly less will be recorded as 183. Any value even slightly less than 185 will be recorded as 184. Thus 184.9 is not the upper boundary of 184. If it was the upper boundary then any value slightly greater would be recorded as 185, but 184.95 is greater than 184.9 and this is recorded as 184. Hence the upper boundary of 184 is 185 since anything slightly less will be recorded as 184. Thus we can construct the table with class boundaries as follows.

Measured length	Boundaries	Frequency
50-51	50-52	7
52-53	52-54	4
54-55	54-56	12
56-57	56-58	5
58-59	58-60	1
60-64	60-65	2
65-69	65-70	2
70-79	70-80	5
80-99	80-100	5
100 -149	100-150	3
150 - 199	150-200	4

5. a)

Iron Intake	Class Boundaries	Frequency	Relative Frequency	Cumulative Frequency
6-8	5.5-8.5	2	0.04	2
9-11	8.5-11.5	7	0.16	9
12-14	11.5-14.5	14	0.31	23
15-17	14.5-17.5	13	0.29	36
18-20	17.5-20.5	8	0.18	44
21-23	20.5-23.5	1	0.02	45

Displays of Data

1. Chironomid larvae (discrete data).

Larvae / Leaf	Frequency	Cumulative Frequency
0	10	10
1	15	25
2	27	52
3	18	70
4	38	108
5	57	165
6	22	187
7	5	192
8	2	194

9	3	197
10	0	197

2. Frequency distribution for iron intake.

Iron Intake (mg)-	Class Boundaries (mg)	Class Length	Frequency	Cumulative Frequency
6-7	5.5-7.5	2	1	1
8-9	7.5-9.5	2	1	2
10-11	9.5-11.5	2	7	9
12-13	11.5-13.5	2	9	18
14-15	13.5-15.5	2	9	27
16-17	15.5-17.5	2	9	36
18-19	17.5-19.5	2	8	44
20-21	19.5-21.5	2	1	45

If the data had been recorded by rounding down to the nearest milligram, all the class boundaries would have been 0.5mg higher. The class boundaries determine where we put the bars and so all the bars would have moved 0.5mg to the right.

3. The data is continuous as we are measuring something (weight).

Birth Weight (kg)	Boundaries	Class Length	Frequency	Height
1.0-1.1	0.95-1.15	0.2	6	6
1.2-1.3	1.15-1.35	0.2	6	6
1.4-1.5	1.35-1.55	0.2	4	4
1.6-1.7	1.55-1.75	0.2	8	8
1.8-1.9	1.75-1.95	0.2	4	4
2.0-2.1	1.95-2.15	0.2	3	3
2.2-2.3	2.15-2.35	0.2	4	4
2.4-2.5	2.35-2.55	0.2	6	6
2.6-2.9	2.55-2.95	0.4	5	2.5
3.0-3.7	2.95-3.75	0.8	4	1

4. Jawbone lengths.

Length (cm)-	Boundaries	Class Length	Frequency	Height
9.0-9.9	8.95-9.95	1.0	2	1
10.0-10.9	9.95-10.95	1.0	5	2.5
11.0-11.4	10.95-11.45	0.5	6	6
11.5-11.9	11.45-11.95	0.5	4	4
12.0-12.4	11.95-12.45	0.5	6	6
12.5-12.9	12.45-12.95	0.5	4	4
13.0-13.9	12.95-13.95	1.0	3	1.5

Averages and Measures of Spread

1. Ant lion pits.

No. pits	Mark c_i	Freq f_i	$f_i\ c_i$	$f_i c_i^2$
0-1	0.5	2	1	0.5
2-3	2.5	3	7.5	18.75
4-5	4.5	6	27	121.5
6-7	6.5	11	71.5	464.75
8-9	8.5	8	68	578
10-11	10.5	4	42	441
12-13	12.5	5	62.5	781.25
14-15	14.5	1	14.5	210.25
		40	294	2616

The mean $\bar{x}$ is given by

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f_i} = \frac{294}{40} = 7.35$$

The variance is given by

$$\mathrm{var}(x) = \frac{\sum_{i=1}^{k} f_i c_i^2}{\sum_{i=1}^{k} f_i} - (\bar{x})^2 = \frac{2616}{40} - (7.35)^2 = 11.3775$$

The standard deviation is $\sqrt{(11.3775)} = 3.37$

2. Mean, median of a small sample.

   a. $Mean = \dfrac{8+9+10+11+12+13+15}{7} = 11.14$

   The median is the (n+1)/2 smallest where n is the number of pieces of data i.e. third smallest. This is 10.

   The variance is $\mathrm{var} = \dfrac{8^2+9^2+10^2+11^2+12^2+13^2+15^2}{7} - 11.14^2 = 5.04$

The (population) standard deviation is $\sqrt{(5.04)}$

b. $$Mean = \frac{8+9+10+11+12+13+15+100}{8} = 22.5$$

Now, there are 6 pieces of data so (n+1)/2 = 4.5. Hence the median is the average of the fourth and fifth smallest which is 11.5. This extra value might have come from an error and this shows that the mean might be strongly affected by an error but the median might not be. Thus the median might be a more appropriate measure to use in this case.

c.

$$Mean = \frac{0+0+0+0+8+9+10+11+12+13+15+100+100+100+100}{15} = 31.867$$

Now there are 15 pieces of data so the median is the (15+1)/2 = 8th smallest. Hence the median is 11. For this data the mean is probably more appropriate since there are repeated values of 100 and 0 and thus they are probably not errors. The 100 is large compared with the other numbers but the median does not reflect this.

3. Jaw Length. Continuous data rounded to the nearest 0.1cm so we need to compute boundaries in the usual way.

Length (cm)	Boundaries (cm)	Mark c_i	Freq f_i	Cumul.Freq	f_i c_i	f_i c_i^2
9.0-9.9	8.95-9.95	9.45	2	2	18.9	178.605
10.0-10.9	9.95-10.95	10.45	5	7	52.25	546.0125
11.0-11.4	10.95-11.45	11.2	6	13	67.2	752.64
11.5-11.9	11.45-11.95	11.7	4	17	46.8	547.56
12.0-12.4	11.95-12.45	12.2	6	23	73.2	893.04
12.5-12.9	12.45-12.95	12.7	4	27	50.8	645.16
13.0-13.9	12.95-13.95	13.45	3	30	40.35	542.7075
			30		349.5	4105.725

The mean $\bar{x}$ is given by

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f_i} = \frac{349.5}{30} = 11.65 \ cm$$

The variance is given by

$$\operatorname{var}(x) = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f} - (\bar{x})^2 = \frac{4105.725}{30} - (11.65)^2 = 1.135$$

The standard deviation is $\sqrt{(1.135)} = 1.07cm$

The median has cumulative frequency (30/2) = 15. The boundary 11.45 has cumulative frequency 13 and the boundary 11.95 has cumulative frequency 17. Hence the median lies between 11.45 and 11.95. Using interpolation

$$Median = 11.45 + \frac{(30/2) - 13}{4} x(11.95 - 11.45) = 11.70cm$$

The lower quartile has cumulative frequency (30/4) = 7.5. The boundary 10.95 has cumulative frequency 7 and the boundary 11.45 has cumulative frequency 13. Hence the lower quartile lies between 10.95 and 11.45. Using interpolation

$$Q1 = 10.95 + \frac{(30/4) - 7}{6} x(11.45 - 10.95) = 10.9917cm$$

The upper quartile has cumulative frequency (90/4) = 22.5. The boundary 12.45 has cumulative frequency 23 and the boundary 12.95 has cumulative frequency 27. Hence the median lies between 12.45 and 12.95. Using interpolation

$$Q_3 = 11.95 + \frac{(90/4) - 17}{6} x(12.45 - 11.95) = 12.8667cm$$

Hence the inter-quartile range is 12.8667 – 10.9917 = 1.875cm

4. Calcium intake. The data is continuous and is obtained by rounding down. Notice what effect this has on the class boundaries. See problem sheet 1 for an explanation. Recall that the class mark is the average of the two boundaries.

Intake-	Boundaries	Midpoint c_i	Freq f_i	Cum. Freq	$f_i c_i$	$f_i c_i^2$
0 –199	0 – 200	100	1	1	100	10000
200 – 399	200 – 400	300	1	2	300	90000
400 –599	400 – 600	500	12	14	6000	3000000
600 – 799	600 – 800	700	16	30	11200	7840000
800 – 999	800 – 1000	900	12	42	10800	9720000
1000 – 1199	1000 – 1200	1100	7	49	7700	8470000
1200 - 1399	1200 - 1400	1300	1	50	1300	1690000
			50		37400	30820000

The mean $\overline{x}$ is given by

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f_i} = \frac{37400}{50} = 748 \ mm$$

The variance is given by

$$\text{var}(x) = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f} - (\overline{x})^2 = \frac{30820000}{50} - (748)^2 = 56896$$

The standard deviation is $\sqrt{(56896)} = 239mm$

The median has cumulative frequency (51/2) = 25.5. The boundary 600 has cumulative frequency 14 and the boundary 800 has cumulative frequency 30. Hence the median lies between 600 and 800. Using interpolation

$$Median = 600 + \frac{(50/2) - 14}{16} x(800 - 600) = 737.5cm$$

The lower quartile has cumulative frequency (50/4) = 12.5. The boundary 400 has cumulative frequency 2 and the boundary 600 has cumulative frequency 14. Hence the lower quartile lies between 400 and 600. Using interpolation

$$Q_1 = 400 + \frac{(50/4) - 2}{12} x(600 - 400) = 575mm$$

The upper quartile has cumulative frequency (150/4) = 37.5. The boundary 800 has cumulative frequency 30 and the boundary 1000 has cumulative frequency 42. Hence the lower quartile lies between 800 and 1000. Using interpolation

$$Q_3 = 800 + \frac{(150/4) - 30}{12} x(1000 - 800) = 925mm$$

Hence the inter-quartile range is 925-575=350

5. Snake litter size. The data is discrete and each class only contains one value so the class mark is that value.

Size	Mark c_i	Freq f_i	Cumul. Freq	f_i c_i	$f_i c_i^2$
5	5	1	1	5	25
6	6	4	5	24	144
7	7	12	17	84	588
8	8	8	25	64	512
9	9	4	29	36	324
10	10	1	30	10	100
		30		223	1693

The mean $\bar{x}$ is given by

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f_i} = \frac{223}{30} = 7.43$$

The variance is given by

$$\operatorname{var}(x) = \frac{\sum_{i=1}^{k} f_i c_i}{\sum_{i=1}^{k} f} - (\bar{x})^2 = \frac{1693}{30} - (7.43)^2 = 1.179$$

The standard deviation is $\sqrt{(1.179)} = 1.09$

The median has cumulative frequency (30)/2 = 15 hence it is the average of the 15th and 16th smallest. By looking at the cumulative frequencies we see that the 15th and 16th smallest are both 7 so the median is 7.

Probability

1. For a child to have the disease it must inherit recessive genes from both parents. The probability that it inherits a recessive gene from the father is 0.5, and so is the probability that it inherits a recessive gene from the mother. We want the probability that both of these happen and since the two things are independent we can use the multiplication rule to get

$$P\ (\text{Child has disease}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

2. The events of getting HEAD and at least 3 on the dice are independent. Hence we can use the multiplication rule.

$$P\ (\text{HEAD and at least 3}) = P\ (\text{HEAD}) \cdot P\ (\text{At least 3}) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

3.
   a. It is tempting to use the multiplication rule here but we cannot. The two events are ``first toss tails'' and ``first toss heads''. Clearly they are not independent because initially each has probability 0.5 of occurring but if we know that the first toss is heads then it is impossible for the first toss to simultaneously be tails so the probabilities are affected and they are not independent. Hence

$$P\ (\text{First tails and first heads}) = 0.$$

   In other words it is impossible for a coin to be both heads and tails at the same time.

   b. P(First toss trail or first toss head)= P(First toss trail)+P(First toss head)=1

   c. The events that the first toss is tails and the second toss is tails are clearly independent since one has no effect on the other. Hence we can use the multiplication rule.

$$P\ (\text{First tails and second tails}) = P\ (\text{First tails}) \cdot P\ (\text{Second tails}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

   d. Since we want to know the probability that the first toss is tails or the second toss is tails it is tempting to use the addition rule. However we may only use the addition rule if the two events are mutually exclusive i.e. they cannot both occur. In this case it is clearly possible for both the first and second tosses to be tails so they are not mutually exclusive. Hence we must go back to first principles. If one or other is tails then the possible ways this can happen are HT, TH, TT i.e. 3 possible ways. There are 4 possible ways that the two coin tosses can land and all are equally likely so

$$P\ (\text{First toss tails or second toss tails}) = \frac{3}{4}.$$

4. 400 workers in total.
    a. If we know the worker is a manager then they are one of the 50 in the bottom line. 20 of these workers for KB. Hence

       P(KB/Manager)=20/50=2/5=0.4

    b. If we know the worker works for KB then they are one of the 170 in the left hand column. 20 of these are managers. Hence

$$P\ (\text{Manager/KB}) = \frac{20}{170} = \frac{2}{17} = 0.118.$$

    c. We need to count how many workers are either in the KB column or in the manager row. Make sure that the ones at the bottom left are not counted twice. There are 200 workers falling into one of these categories and so.

       P(Either KB or Manger)=p(KB)+p(Manager)

       = 170/400+50/400- 2/400=0.5

5. To get from A to C we must go from A to B and then from B to C. Need at least one open road from A to B and at least one open road from B to C. All roads are independent so we can use the multiplication rule. We will first work out the probability that we can get from A to B. Using the ``not'' rule we can say that the probability that we can get from A to B can obtained by subracting the probability that we can't get from A to B away from 1. The probability that we can't get from A to B is just the probability that both roads are blocked.

$$P\ (\text{Both roads blocked from A to B}) = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100}.$$

$$P\ (\text{At least 1 road open A to B}) = 1 - \frac{1}{100} = \frac{99}{100}.$$

Similarly

$$P\ (\text{At least 1 road open B to C}) = 1 - \frac{1}{100} = \frac{99}{100}.$$

The events that at least one road is open from A to B and that at least one road is open from B to C are independent. So we can use the multiplication rule.

$$P\ (\text{Get from A to C}) = P\ (\text{Get from A to B}) \cdot P\ (\text{Get from B to C}) = \frac{99}{100} \cdot \frac{99}{100} = 0.9801.$$

6.

a. There are five balls, three of which are red. Each of the five is equally likely to be chosen first. Hence

$$P\ (\text{First red}) = \frac{3}{5}.$$

b. There are five balls, three of which are red. Each of the five is equally likely to be chosen second. Hence

$$P\ (\text{Second red}) = \frac{3}{5}.$$

c. If you are not convinced and think that we should be considering the colour of the first ball then first notice that the question says nothing about the colour of the first ball, we have not been told it is red or blue. Maybe you think that after we take the first ball out we will know what colour it is and this will affect the chances of the second ball being red. This is true but the first ball could be either red or blue and so in the end the probability works out at being 3/5. One way to think about it is that we could take out the first ball and not look at it. Now each ball has equal chance of being the one we have taken out and not looked at and the one we are about to take out. Here we need to use the conditional equation.

$$P(\text{A and B}) = P(\text{A/B})\, P(\text{B}),$$

Where A is the second ball being red and B is the first ball being red. That is

$$P(\text{Both Red}) = P(\text{Second Red/First Red})\, P(\text{First Red}).$$

If we are told that the first ball is red then there are 4 balls left only two of which are red and hence the probability that the second ball is red is 2/4. Notice this is difference from part b where we are told nothing. Substituting this in we get

$$P(\text{Both Red}) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}.$$

First work out the probability that both balls are blue. We do a similar calculation to the one above and get

$$P(\text{Both Blue}) = P(\text{Second Blue/First Blue})\, P(\text{First Blue})$$

$$= \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{10}.$$

Using the addition rule and the previous parts we see that

$$P(\text{Both same colour}) = \frac{3}{10} + \frac{1}{10} = \frac{2}{5}.$$

d. Now using the not rule we see that

$$P(\text{One red, one blue}) = 1 - P(\text{Both same colour}) = 1 - \frac{2}{5} = \frac{3}{5}.$$

7.

a. For a child to have the recessive characteristic they must inherit recessive genes from both parents. Consider the child's mother. Let her genes be G_1G_2. The child has the recessive characteristic if it inherits G_1 and G_1 is recessive or it inherits G_2 and G_2 is recessive.

$$P\ (\text{Inherits } G_1 \text{ and it is recessive}) = \frac{1}{2}\ \frac{3}{10} = \frac{3}{20}\ .$$

Here we have used independence. Now using the addition rule.

$$P\ (\text{Inherits recessive from M}) \ = \ P\ (\text{Inherits } G_1 \text{ and it is rec}) + P\ (\text{Inherits } G_2 \text{ and it is rec})$$

$$= \frac{3}{20} + \frac{3}{20} = \frac{3}{10}\ .$$

Now using independence

$$P\ (\text{Child has rec char}) \ = \ P\ (\text{Inherits rec gene from M}) \cdot P\ (\text{Inherits rec gene from F})$$

$$= \frac{3}{10} \cdot \frac{3}{10} = \frac{9}{100}\ .$$

First find the probability that the mother has the recessive characteristic given that she has passed on a recessive gene. The probability that she has passed on a recessive gene is

$$\frac{3}{10} \cdot \frac{3}{10} + \frac{1}{2}\ \frac{3}{10}\ \frac{7}{10} + \frac{1}{2}\ \frac{3}{10}\ \frac{7}{10} = \frac{3}{10}\ .$$

Hence the probability that she has the recessive characteristic given that she has passed on the recessive gene is

$$\frac{9}{100} / \frac{3}{10} = \frac{3}{10}\ .$$

Now the probability that both parents have the recessive characteristic given that they have both passed on the recessive gene is

$$\frac{3}{10} \cdot \frac{3}{10} = \frac{9}{100}\ .$$

This is the answer we want since both parents passing on the recessive gene is equivalent to the child having the recessive characteristic.

8. Suppose we have chosen door 1 and let's look at what might happen next. Firstly the car may be behind door 3, in which case the host will open door 2. This happens with probability 1/3. Secondly is the case when the car is behind door 2, in which case the host must open door 3. Again this happens with probability 1/3. Now things get a bit more complicated. Suppose the car is behind door 1. This happens with probability 1/3. However the host has a 50:50 choice of two doors to open to show us a goat. So the probability that the car is behind the first door and the host opens the second door is 1/6 whereas the probability that the car is behind the first door and the host opens the third door is 1/6. Suppose in actual fact the host opens the second door. We need to calculate

$$P\ (\text{Car behind door 1 / Host opens door 2}).$$

Using the conditional rule, we see that

$$P\ (\text{Car behind door 1/ Host opens door 2}) = \frac{P\ (\text{Car behind door 1 and host opens door 2})}{P\ (\text{Host opens door 2})}.$$

If we look at the discussion above we see that the probability that the car is behind door 1 and the host opens door 2 is 1/6 whereas if we add the probabilities for the two cases where the host opens door 2 we see that the probability that the host opens door 2 is 1/2. Hence the probability that the car is behind door 1 once we see the host open door 2 is

$$\frac{1/6}{1/2} = \frac{1}{3}.$$

Thus we are twice as likely to find the car if we switch choices of door.

Binomial, Poisson and Normal Distributions

1. Let X be the number of children who have the disease. Then X has the Binomial distribution with parameters n = 6 and p = 0. 25.

$$P(X = x) = C_x^6 (0.25)^x (0.75)^{6-x}, x = 0,1,2,...,6$$

   a. We use the Binomial formula.

$$P(x=2)=15.(0.25)^2 (0.75)^4 =0.297$$

   b. We need to find $P(X \leq 1)$ and we begin by observing that

$$P(X \leq 1)=P(x=0)+P(x=1)$$

   Using the Binomial formula we get

$$P(x \leq 1) = \frac{6!}{0!(6-0)!}(0.25)^0(0.75)^6 + \frac{6!}{1!(6-1)!}(0.25)^1(0.75)^5 = 0.534$$

   c. The expected number of children is equal to **1.5**

2. The answers to the first part are obtained just by looking them up in the table in the standard way.
   a. $P(Z \leq 1) = 0.8413$.
   b. $P(Z \geq 0.5) = 1-P(Z \leq 0.5) = 1-0.6915 = 0.3085$.
   c.

$$P(-1.5 \leq Z \leq 2)=P(Z \leq 2)-P(Z \leq -1.5)=0.9772-0.0668=0.9104$$

The second part of the question is completely different to the first part. It essentially consists of doing the first part backwards. We look the number up in the body of the table and find the z number that would give it.

   a. If $Pr(Z \leq z) = 0.5$ then if we look up 0.5 in the body of the table we see that z = 0.
   b. If we look up 0.95 in the body of the table we see that z lies between 1.64 and 1.65.

3. Let X be the brain weight of a male elephant. Then X $\dot{\sim}$ N(1.4,0.11).

$$P(1.5 \le x \le 1.6) = P(x \le 1.6) - P(x \le 1.5)$$

$$= P(Z \le \frac{1.6 - 1.4}{\sqrt{0.11}}) - P(Z \le \frac{1.5 - 1.4}{\sqrt{0.11}}) = 0.7257 - 0.6179 = 0.1078$$

Let

$$P(1.2 \le x \le 1.3) = P(x \le 1.3) - P(x \le 1.2)$$

$$= P(Z \le \frac{1.2 - 1.4}{\sqrt{0.11}}) - P(Z \le \frac{1.3 - 1.4}{\sqrt{0.11}}) = 0.3821 - 0.2743 = 0.1078$$

This is the same answer as in the previous part. The reason for this is that the two ranges 1.2-1.3 and 1.5-1.6 are symmetric about the mean 1.4 and since the Normal distribution is symmetric about the mean we would expect the answers to be the same.

4. Let X be the production of 11-KT. Then X $\dot{\sim}$ N(100,10).

$$P(100 \le x \le 101) = P(x \le 101) - P(x \le 100)$$

$$= P(Z \le \frac{101 - 100}{\sqrt{10}}) - P(Z \le \frac{100 - 100}{\sqrt{10}}) = 0.6255 - 0.5 = 0.1255$$

Finding the value that is exceeded 1% of the time means finding the value x such that

$$P(X \le x) = 0.99$$

We must first find the z value that is exceeded 1% of the time, that is the value z such that

$$P(Z \le z) = 0.99$$

Looking in the main body of the table we see that z = 2.33. We know that

$$Z = \frac{x - 100}{\sqrt{10}}$$

We need to substitute z = 2.33 into this equation and find x.

$$2.33 = \frac{x - 100}{\sqrt{10}} \rightarrow x = 107.4$$

Thus the 11-KT level that is exceeded only 1% of the time is 107.4pg/mg.

5. dLet X be the number of ants' nests in a 10m quadrate. We know that X has the Poisson distribution with parameter 2.3. The formula for the Poisson distribution is

$$P(X = x) = \frac{(2.3)^x}{x!} e^{-2.3}$$

Thus

$$P(X = 4) = \frac{(2.3)^4}{4!} e^{-2.3} = 0.117$$

For the second part

$$P(x \geq 3)=1-P(x<3)=1-[P(x=0)+P(x=1)+P(x=2)]=0.404$$

6. The distribution is Binomial. The probability that a biologist gets trapped in the lift on any one day is 0.0001. The biologist goes to work 2500 times. Hence the distribution is B(2500,0.0001). Thus the expected number of times that the biologist gets stuck in the lift is 2500 ·0.0001 = 0.25.

   Assuming the biologist takes the stairs the distribution of the number of times that they meet the head of department in a year is B(250,0.1). We try to approximate this using the Normal distribution. If X ~B(250,0.1) then

$$E(x)=(250)(0.1)=25 \text{ and } var(x)=(250)(0.1)(1-0.1)=22.5$$

   Thus we approximate X by N(25,22.5). This is appropriate since np = 25 > 5 and np(1-p) = 225 > 5. Since we are approximating a discrete distribution using a continuous one we use the boundaries and find $P(19.5 \leq X \leq 25.5)$ using N(25,22.5).

$$P(19.5 \leq x \leq 25.5)=P(x \leq 25.5)-P(x \leq 19.5)$$

$$= P(Z \leq \frac{25.5-25}{\sqrt{22.5}}) - P(Z \leq \frac{19.5-25}{\sqrt{22.5}}) = 0.5438-0.1230=0.4208$$

7. If X has the Poisson distribution with parameter 100 then it is difficult to calculate probabilities for it - try finding e^{100}. We approximate X by N(100,100) and use the boundaries to find the probabilities.

$$P(98.5 \leq x \leq 101.5)=P(x \leq 101.5)-P(x \leq 98.5)$$

$$= P(Z \leq \frac{101.5-100}{\sqrt{100}}) - P(Z \leq \frac{98.5-100}{\sqrt{100}}) = 0.5596-0.4404=0.1192$$