

# Exercises Sheet 1

## Types of Statistical Data and Display of Data

1. For each of the following decide whether the data is continuous, discrete or categorical.
  - a. The number of hours of operation of a number of 100W light bulbs.
  - b. The number of current-account balances checked by a firm of auditors each year.
  - c. The cost of bed and breakfast in three-star London hotels.
  - d. The occupation of adult females.
  - e. The number of failures in each week of a year of a large computer system.
  - f. The down time of a computer system in each week of a year.
  - g. The grades attained by candidates taking A-level mathematics.
  - h. The reaction time of rats to a stimulus.
  
2. The number of eggs produced in a single brood by 30 female iguanas was as follows.

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 33 | 50 | 46 | 33 | 53 | 57 | 60 | 40 | 31 | 38 |
| 28 | 44 | 51 | 37 | 42 | 45 | 49 | 55 | 52 | 39 |
| 34 | 37 | 32 | 41 | 55 | 43 | 53 | 52 | 42 | 28 |

- a. Is the data discrete or continuous?
  - b. Group the data into an appropriate number of equally sized classes and calculate the frequencies, relative frequencies and cumulative frequencies.
  
3. The arm spans of 30 adult male employees of a certain firm were measured. They were rounded to the nearest centimeter and recorded.

|     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|
| 173 | 186 | 180 | 170 | 186 | 179 |
| 183 | 172 | 187 | 169 | 184 | 183 |
| 178 | 171 | 184 | 174 | 181 | 179 |
| 186 | 178 | 177 | 174 | 170 | 169 |
| 168 | 172 | 180 | 181 | 179 | 177 |

- a. Draw up a frequency table with 10 classes and then with 5 classes. Include the class boundaries, relative frequencies and cumulative frequencies.
  - b. Repeat the first part with 5 classes but now suppose that the lengths are recorded by *rounding down* to the nearest centimetre. So span 184.6cm would be recorded as 184cm. What difference does this make?

4. The following data showing the length of 50 bluegill fish is already grouped into classes.

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

- Calculate the class boundaries assuming the data is recorded by rounding to the nearest mm.
  - Calculate the class boundaries assuming the data is recorded by rounding down to the nearest mm.
  - Calculate the relative and cumulative frequencies.
5. The following data shows the 24-hour iron intake recorded to the nearest milligram for a sample of 45 women.

|    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|
| 8  | 10 | 11 | 14 | 14 | 18 | 12 | 13 | 16 |
| 18 | 17 | 10 | 13 | 14 | 17 | 16 | 12 | 13 |
| 21 | 17 | 16 | 14 | 15 | 12 | 13 | 18 | 13 |
| 17 | 13 | 19 | 6  | 15 | 16 | 15 | 17 | 15 |
| 11 | 10 | 10 | 10 | 14 | 18 | 18 | 18 | 19 |

6. Group the data into a suitable number of classes and construct a frequency table with class boundaries, relative frequencies and cumulative frequencies.

## Displays of Data

1. Some species of chironomid larvae inhabit the leaves of pitcher plants. The number of larvae per leaf in a random sample of 197 leaves was as follows.

| Larvae / Leaf | Number of leaves with<br>A particular number of larvae |
|---------------|--|
| 0             | 10   |
| 1             | 15   |
| 2             | 27   |
| 3             | 18   |
| 4             | 38   |
| 5             | 57   |
| 6             | 22   |
| 7             | 5  |
| 8             | 2  |
| 9             | 3  |
| 10            | 0  |

- a) Display this information in a bar chart and a cumulative frequency polygon. Repeat using SPSS.
2. The following is the 24-hour iron intake recorded to the nearest milligram for a sample of 45 women.

| Iron Intake (mg) | Frequency |
|------------------|-----------|
| 6-7              | 1         |
| 8-9              | 1         |
| 10-11            | 7         |
| 12-13            | 9         |
| 14-15            | 9         |
| 16-17            | 9         |
| 18-19            | 8         |
| 20-21            | 1         |

- a) Draw a histogram and cumulative frequency polygon to depict this data. What difference would it make if the data was recorded by rounding down to the nearest milligram?
3. The following is the set of birth weights of children with severe idiopathic respiratory disease syndrome measured to the nearest 100g.

| Birth Weight (kg) | Frequency |
|-------------------|-----------|
| 1.0-1.1           | 6         |
| 1.2-1.3           | 6         |
| 1.4-1.5           | 4         |
| 1.6-1.7           | 8         |
| 1.8-1.9           | 4         |
| 2.0-2.1           | 3         |
| 2.2-2.3           | 4         |
| 2.4-2.5           | 6         |
| 2.6-2.9           | 5         |
| 3.0-3.7           | 4         |

a) Is the data continuous or discrete?. Draw a histogram to display this data.

4. The bill lengths of 42 belted kingfishers were measured to the nearest millimetre.

| Measured Length | Boundaries | Class Length | Frequency |
|-----------------|------------|--------------|-----------|
| 45-49           | 44.5-49.5  | 5            | 3         |
| 50-51           | 49.5-51.5  | 2            | 5         |
| 52-53           | 51.5-53.5  | 2            | 9         |
| 54-55           | 53.5-55.5  | 2            | 8         |
| 56-57           | 55.5-57.5  | 2            | 7         |
| 58-59           | 57.5-59.5  | 2            | 5         |
| 60-64           | 59.5-64.5  | 5            | 3         |
| 65-74           | 64.5-74.5  | 10           | 2         |

a) Draw a histogram to display this data

5. The following data shows the jawbone lengths measured to the nearest 0.1 cm of 30 kangaroos of the species *Macropus Giganteus*.

| Length    | Frequency |
|-----------|-----------|
| 9.0-9.9   | 2         |
| 10.0-10.9 | 5         |
| 11.0-11.4 | 6         |
| 11.5-11.9 | 4         |
| 12.0-12.4 | 6         |
| 12.5-12.9 | 4         |
| 13.0-13.9 | 3         |

a) Draw a histogram to display this data.

6. Use the data from problem sheet 1 questions 2 and 3 to use SPSS to draw bar charts, cumulative frequency polygons and histograms.

## Averages and Measures of Spread

1. A survey was carried out to determine the number of ant lion pits in 40 one-metre-square quadrates.

| No. Pits | Frequency |
|----------|-----------|
| 0-1      | 2         |
| 2-3      | 3         |
| 4-5      | 6         |
| 6-7      | 11        |
| 8-9      | 8         |
| 10-11    | 4         |
| 12-13    | 5         |
| 14-15    | 1         |

- a) Find the mean and standard deviation of this data.
2. In an experiment the following data values were recorded: 8,9,10,11,12,13,15.
- Find the mean, median and population standard deviation of this data.
  - Now suppose there was a further value of 100 recorded. What difference does this make to the mean and the median? How might this data-set with the one value of 100 arise in an experiment? Comment on the implications this has for the usefulness of the mean and the median.
  - Now suppose that in addition to the original 5 values, there were 4 more values of 100 and 4 values of 0. What difference does this make to the mean and the median? Comment again on your results.
3. The following data shows the jawbone lengths measured to the nearest 0.1 cm of 30 kangaroos of the species *Macropus Giganteus*.

| Length    | Frequency |
|-----------|-----------|
| 9.0-9.9   | 2         |
| 10.0-10.9 | 5         |
| 11.0-11.4 | 6         |
| 11.5-11.9 | 4         |
| 12.0-12.4 | 6         |
| 12.5-12.9 | 4         |
| 13.0-13.9 | 3         |

- a) Calculate the mean, median, standard deviation and inter-quartile range for this data.?

4. The recommended daily allowance of calcium for adults is 800mg. A nutritionist suspects that people living in a particular area have a low intake of calcium. To test out her suspicion she measures the calcium intake of 50 people from the area. The data is recorded rounded down to the nearest mg.

| Intake (mg) | Frequency |
|-------------|-----------|
| 0-199       | 1         |
| 200-399     | 1         |
| 400-599     | 12        |
| 600-799     | 16        |
| 800-999     | 12        |
| 1000-1199   | 7         |
| 1200-1399   | 1         |

- a) Compute the mean, median, standard deviation, and interquartile range for this sample. Is the class mark the average of the largest and smallest measured values in each class? Explain why or why not.
5. The following shows the litter size of 30 garter snakes. Find the mean, median, and standard deviation of this data.

| Litter Size | Frequency |
|-------------|-----------|
| 5           | 1         |
| 6           | 4         |
| 7           | 12        |
| 8           | 8         |
| 9           | 4         |
| 10          | 1         |

- a) Compute the mean, median, standard deviation
-

# Probability

1. A couple are heterozygous for sickle cell anaemia. What is the probability that their first child has sickle cell anaemia?
2. A fair coin and a fair dice are thrown. What is the probability that we get a HEAD and at least 3 on the dice?
3. A coin is tossed twice.
  - a.  $\Pr(\text{First toss tails and First toss heads})$ ,
  - b.  $\Pr(\text{First toss tails or First toss heads})$ ,
  - c.  $\Pr(\text{First toss tails and Second toss tails})$ ,
  - d.  $\Pr(\text{First toss tails or Second toss tails})$ .

4. The following shows the split of workers in two companies into manual, clerical and management.

|            | KB  | McDougal's |
|------------|-----|------------|
| Manual     | 100 | 150        |
| Clerical   | 50  | 50         |
| Management | 20  | 30         |

- a. If a given worker is a manager, what is the probability they work for KB?
  - b. If a given worker works for KB what is the probability that they are a manager?
  - c. What is the probability that a given worker either works for KB or is a manager
5. There are two roads from A to B and two roads from B to C. Each of the four roads has independently, probability  $1/10$  of being blocked by snow. What is the probability that I can drive from A to C?
6. We have an urn containing five balls, three red and two blue. We choose a ball and then without replacing it choose a second. Find
  - a. The probability that the first ball is red.
  - b. The probability that the second ball is red.
  - c. The probability that both balls are red.

- d. The probability that one ball is red and the other is blue.
7. \* Harder question!! This question uses GCSE genetics. Suppose within a population 70% of the genes for one particular characteristic are of the dominant type and 30% are of the recessive type. Suppose we take two adults. Assume their genes are chosen independently at random but with the given probabilities.
- What is the probability that their first child shows the recessive characteristic?
  - Given that their only child shows the recessive characteristic, what is the probability that both parents show the recessive characteristic?
8. \*\* Much harder question!! At the end of a game show the winning contestant is shown three doors. Behind one door there is a car and behind the other two are goats. The objective of the contestant is to try to guess the door, which hides the car. The contestant begins by choosing one of the doors. The game show host knows where the car is and now opens one of the two doors not chosen by the contestant and reveals a goat. He now offers the contestant the chance to change the choice of doors i.e. the contestant can stick with the original choice or pick the other door that was not opened by the game show host. What should the contestant do? Does it matter? Why?
-

## Binomial, Poisson and Normal Distributions

1. In couples where each person is heterozygous for the sickle-cell gene, there is probability 0.25 that any child of the couple will actually have the disease. In a randomly selected family of 6 children where both parents have the disease,
  - a. What is the probability that 2 children have the disease?
  - b. What is the probability that at most 1 child has the disease?
  - c. What is the expected number of children in the family to have the disease?
  
2. Let  $Z$  have the standard Normal distribution. What is the probability that
  - a.  $Z \leq 1$ ?
  - b.  $Z \geq 0.5$ ?
  - c.  $-1.5 \leq Z \leq 2$ ?

What is the value  $z$  such that

- d.  $P(Z \leq z) = 0.5$ ?
  - e.  $P(Z \leq z) = 0.95$ ?
  
3. Male elephants are known to have brain weights that are approximately Normally distributed with mean 1.4 kg and variance 0.11 kg. Find the approximate proportion of male elephants with brain weight between 1.5 kg and 1.6 kg. Now find the approximate proportion with brain weight between 1.2kg and 1.3kg. Explain the link between the two answers.
  
4. An experiment was carried out to investigate the production of a steroid hormone 11-ketotestosterone (11-KT). The amount of 11-KT produced depends on the concentration of gonadotropic hormone and it is known that subjected to a concentration of 15 ng/ml of gonadotropic hormone, the production of 11-KT is Normally distributed with mean 100pg/mg of tissue and variance 10. Find the probability that the 11-KT production is between 100pg/mg and 101pg/mg. What value of hormone production is only exceeded 1% of the time.

5. A species of ant, *Messor Wasmani* inhabits northern Greece. Within a particular wooded area, it is known that the number of nests of *Messor Wasmani* in a 10m quadrat follows the Poisson distribution with mean 2.3. Find the probability that a randomly chosen quadrat has 4 ants' nests and the probability that a randomly chosen quadrat has at least 3 ants' nests.
  
  6. A university administrator assures a biologist that they have only 1 in 10,000 chance of being trapped in a lift on any given day. Assuming the biologist goes to work 5 days a week, 50 weeks of the year and for 10 years what is the distribution of the number of times the biologist gets trapped in the lift? What is the expected number of times that the biologist will be trapped in the lift? The biologist now decides that the risk of going up in the lift is too large and decides to take the stairs. The probability that they meet their head of department on the stairs and therefore gets lots of extra work to do is 0.1. Using a Normal approximation find the probability that the biologist meets their head of department between 20 and 25 times in a year.
  
  7. Suppose the variable  $X$  has the Poisson distribution with parameter 100. Using a Normal approximation find the probability that  $X$  lies between 99 and 101 inclusive.
-