DUX College

Phone: (02) 8007 6824
Email: info@dc.edu.au
Web: dc.edu.au

2018 HIGHER SCHOOL CERTIFICATE
COURSE MATERIALS

Year 9 Headstart Mathematics

Statistics

Term 1 – Week 2

Name …….

Class day and time ……

Teacher name ……...

Term 1 – Week 2 – Theory

Finding the mean from a frequency distribution table

To find the mean from a frequency distribution table, **an extra column (fx) is added** to a frequency distribution table consisting of the **number of outcomes (x) multiplied by the frequency (f)**. This fx column is important as we need to consider any values that may have been duplicated. Once the fx column has been formed, we can calculate the total of all scores ($sum\ of\ all\ fx\ values$) and the total number of scores ($sum\ of\ all\ f\ values$), allowing us to calculate the mean by using the following formula:

$$Mean = \bar{x} = \frac{Total\ of\ all\ scores}{Total\ of\ the\ number\ of\ scores} = \frac{Sum\ of\ all\ fx\ values}{Sum\ of\ all\ f\ values}$$

Example:

Julie purchased a bag of oranges and wanted to find the average weight (in grams) of each orange. Determine the average weight of each orange by using the following frequency distribution table to the nearest gram:

Weight of individual orange (g)	f
375	1
400	1
425	2
450	3
475	3
500	2
525	1

Solution:

Firstly, we need to **calculate the values in the fx column** by multiplying the weight of the individual oranges (x) by their amount (f). Note: Weight is in grams.

Weight of each individual orange (x)	f	fx
375	1	$= 375 \times 1 = 375$
400	1	$= 400 \times 1 = 400$
425	2	$= 425 \times 2 = 850$
450	3	$= 450 \times 3 = 1350$
475	3	$= 475 \times 3 = 1425$
500	2	$= 500 \times 2 = 1000$
525	1	$= 525 \times 1 = 525$

Since we can now **calculate the total of all scores** and the **total of the number of scores**, we can determine the mean.

$$\therefore Mean = \bar{x} = \frac{Total\ of\ all\ scores}{Total\ of\ the\ number\ of\ scores}$$
$$= \frac{375 + 400 + 850 + 1350 + 1425 + 1000 + 525}{1 + 1 + 2 + 3 + 3 + 2 + 1}$$
$$= 456\ grams$$

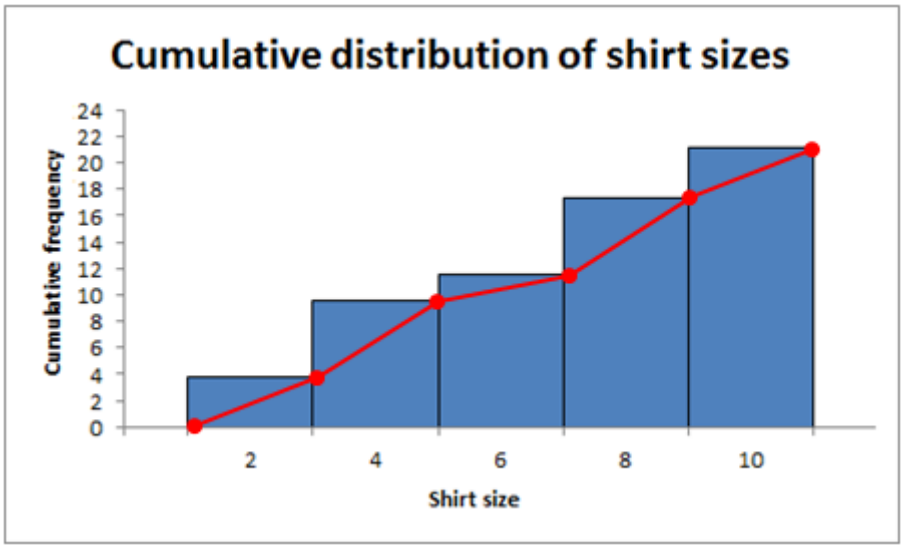
Finding the mean from an ogive

The simplest way to find the mean from an ogive involves **determining the frequency of each unique value**, allowing us to calculate fx values and consequently the mean. The following formula is a handy way to calculate the frequency of each unique value:

$$\boldsymbol{f = c.f.\ of\ current\ value - c.f.\ of\ previous\ value}$$

Example:

John decided to buy his grandchildren t-shirts for Christmas and has recorded the shirt sizes of each grandchild to make sure that the shirts fit properly. Determine the mean shirt size from the below ogive John has constructed with the information he recorded:



Solution:

To determine the mean, we will **construct a cumulative frequency distribution table** using the information from the ogive to aid us in the calculation. We can then use the table to find the frequency by using the formula $f = c.f.of\ current\ value - c.f.of\ previous\ value$, allowing for the fx values and therefore the mean to be calculated (similarly to the first example in this workbook).

Shirt size (x)	$c.f.$	f	fx
2	4	$= 4 - 0 = 4$	8
4	10	$= 10 - 4 = 6$	24
6	12	$= 12 - 10 = 2$	12
8	18	$= 18 - 12 = 6$	48
10	22	$= 22 - 18 = 4$	40

$$\therefore Mean = \bar{x} = \frac{Total\ of\ all\ scores}{Total\ of\ the\ number\ of\ scores}$$

$$= \frac{8 + 24 + 12 + 48 + 40}{4 + 6 + 2 + 6 + 4}$$

$$= 6$$

Therefore, the mean shirt size is 6.

Note: The total of the number of scores for an ogive is equal to the $c.f.$ of the final value.

Discrete and continuous data

Data is considered to be **discrete** if only a **finite number of values** are possible and in many cases, involves integers. For example, imagine counting 10 oranges; 1 orange, 2 oranges, 3 oranges and so on. As you can see, we would not consider anything but an orange as a unit, i.e. half an orange is not considered as a whole orange by rounding up. In that sense, a discrete unit is indivisible and if a unit is divided, it will not be considered as a unit for counting.

Continuous data makes up the rest of the data, where an **infinite number of values** are possible and is usually associated with physical measurement. Using the oranges example from above, continuous data **considers fractions** of oranges as data where there are no limits to how small these fractions can be.

Example:

Decide which type of data (continuous or discrete) should be used to describe the following situations:

a) Counting and recording the amount of whole circle chocolate pieces in a box of circle chocolates
b) Counting and recording the amount of whole circle chocolate pieces in a box of circle chocolates where each and every chocolate piece has been cut in half
c) Counting and recording the amount of whole chocolate pieces in a box of circle chocolates where each and every chocolate piece has been cut in half
d) Determining and recording the heights of a class of students

Solution:

a) The data obtained by counting the amount of whole circle chocolate pieces is discrete as only whole circle pieces of chocolate are considered, i.e. fractions of a piece of chocolate are not considered; 1 whole circle chocolate, 2, 3 and so on.
b) The data is considered discrete similarly to a), but since we know that each and every chocolate piece has been cut in half, we can identify that the number of whole circle chocolate pieces is equal to 0.
c) The data again is considered to be discrete similarly to b). In this case however, any whole pieces (including pieces that have been cut) are considered, i.e. 1 whole chocolate, 2, 3 and so on.
d) The data obtained is considered to be continuous as there are endless possibilities in the heights of students, i.e. 150cm, 150.1cm, 150.11cm, 150.111cm and so on are all possible values in continuous data.

Grouping Data

Individual outcomes in a set of values may be **arranged into groups or classes** making them easier and more convenient to analyse, as well as sometimes allowing the use of continuous data. For example, a national age survey was conducted to see the variations in frequency between age groups (with age intervals of 10 years) where people between the age of 1 and 10 are classed in the 1-10 age group, the people between the age of 11 and 20 are classed in the 11-20 age group and so on. Note: The first number is the lower class limit and the second number is the upper class limit.

The class centre (c.c.) is a value that lies at a point such that it splits the class interval directly in half. The class centre can be calculated by the following formula:

$$Class\ centre = \frac{Upper\ class\ limit - Lower\ class\ limit}{2}$$

Note: Think of it as the average of the highest and lowest values in a class.

The modal class is the class with the highest frequency.

The median class is the class containing the middle value, i.e. the class that contains the median.

The **mean** can be calculated when data is grouped by finding the total of all scores and the total of the number of scores, similarly to finding the mean in a frequency distribution table and an ogive. However, the key difference for grouped data is that the total of all scores is calculated by multiplying the class centre of each class by the frequency of each class instead of multiplying each unique value by its frequency. The following formula represents this mathematically:

$$Mean\ for\ grouped\ data = \frac{Total\ of\ all\ scores}{Total\ of\ the\ number\ of scores} = \frac{Sum\ of\ all\ f \times c.c.\ values}{Sum\ of\ all\ f\ values}$$

Example:

A local community decided to run an Easter egg hunt where the winner receives a lifetime supply of chocolate by collecting the most eggs within the allocated time. A bonus egg was given to each contestant so that every person does not leave empty handed; the bonus egg contributes to the final egg count. Using the following data:

12	24	37	47	67	82	91	57	78	5	31	85	69
17	64	84	72	67	94	83	73	82	73	28	18	37
56	48	39	75	93	73	62	18	47	27	47	28	33
4	16	7	42	17	32	28	94	74	38	29	1	8

a) Group the egg count into classes
b) Calculate the class centres
c) Draw a tally
d) Determine $f, f \times c.c., c.f$.
e) Determine the modal class, median class and mean (to the nearest whole number)
f) Sketch a frequency histogram and polygon
g) Sketch a cumulative frequency histogram and polygon

Solution:

a), b), c) and d)

For this example, we will be using classes with intervals of 10 eggs.

Egg count classes	Class centre (c.c.)	Tally	f	$f \times c.c.$	$c.f.$
1-10	$= \frac{1+10}{2} = 5.5$	𝍸	5	27.5	5
11-20	$= \frac{11+20}{2} = 15.5$	𝍸 |	6	93	11
21-30	$= \frac{21+30}{2} = 25.5$	𝍸 |	6	153	17
31-40	$= \frac{31+40}{2} = 35.5$	𝍸 ||	7	248.5	24
41-50	$= \frac{41+50}{2} = 45.5$	𝍸	5	227.5	29
51-60	$= \frac{51+60}{2} = 55.5$	||	2	111	31
61-70	$= \frac{61+70}{2} = 65.5$	𝍸	5	327.5	36
71-80	$= \frac{71+80}{2} = 75.5$	𝍸 ||	7	528.5	43
81-90	$= \frac{81+90}{2} = 85.5$	𝍸	5	27.5	48
91-100	$= \frac{91+100}{2} = 95.5$	||||	4	22	52

e) There are two modal classes in this example; the egg count classes with 31-40 eggs and 71-80 eggs (both have the highest frequency of 7 people).

The median class has a score that lies between the 26th and the 27th values. Therefore the median class is the 41-50 egg count class as the 26th and 27th values reside in this class.

Using the mean formula from the theory section:

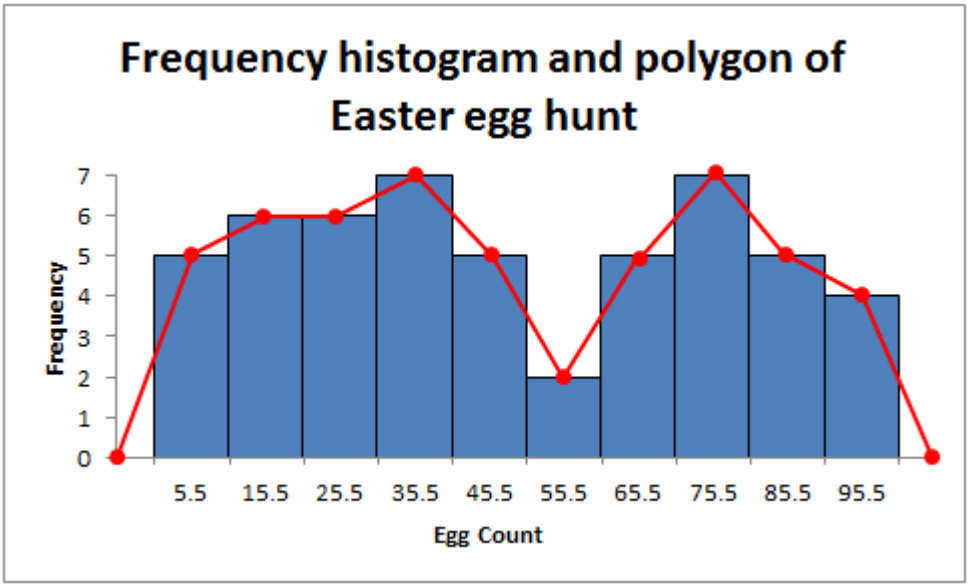
$$\bar{x} = \frac{Sum\ of\ all\ f \times c.c.\ values}{Sum\ of\ all\ f\ values}$$

$$= \frac{27.5 + 93 + 153 + 248.5 + 227.5 + 111 + 327.5 + 528.5 + 27.5 + 22}{52}$$

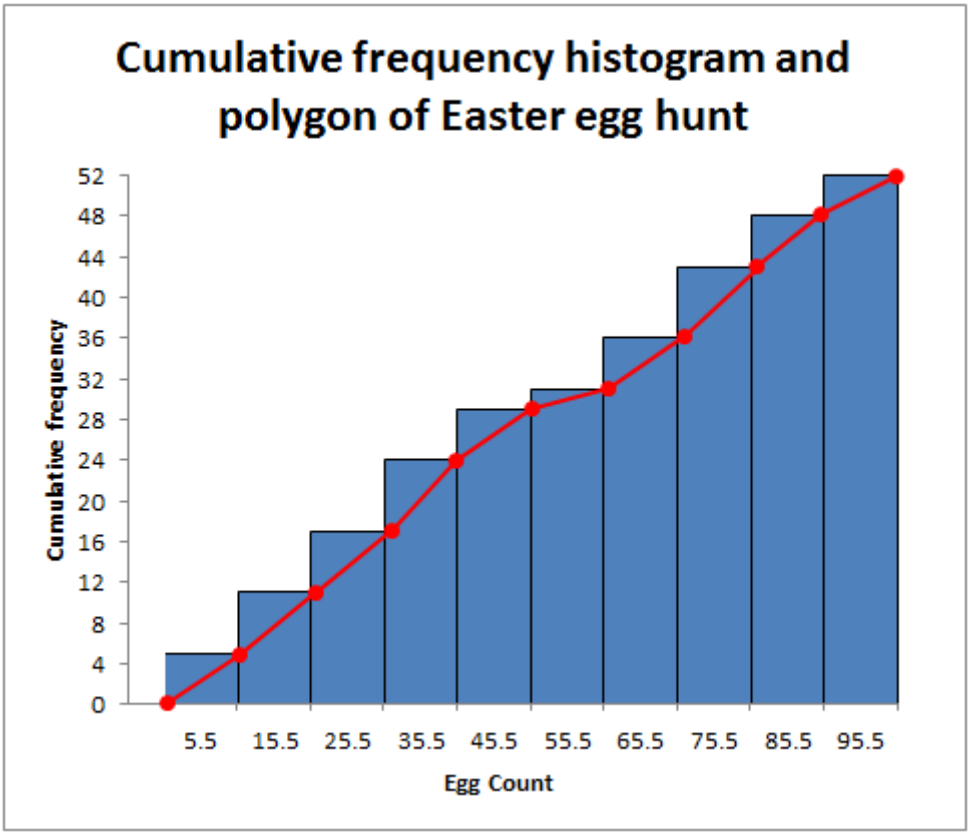
$$= 33.96$$

$= 34$ Eggs (to the nearest whole number)

f) For frequency histograms and polygons, we use the class centre instead of the class interval to indicate the columns.



g) Similarly for cumulative frequency histograms and polygons, we use the class centre instead of the class interval to indicate the columns.



Term 1 – Week 2 – Homework

Finding the mean from a frequency distribution table and an ogive

1. Lisa, a shoe fanatic, decided to go shoe hunting in a shopping centre to record the sizes of all the red shoes she could find. Determine the mean shoe size by using the following frequency distribution table:

Shoe size	f
2	1
3	1
5	2
6	4
7	6
8	7
9	3
10	1

2. Determine the average weight of fish (in grams) by using the following frequency distribution table of the weights of fish caught by a fisherman:

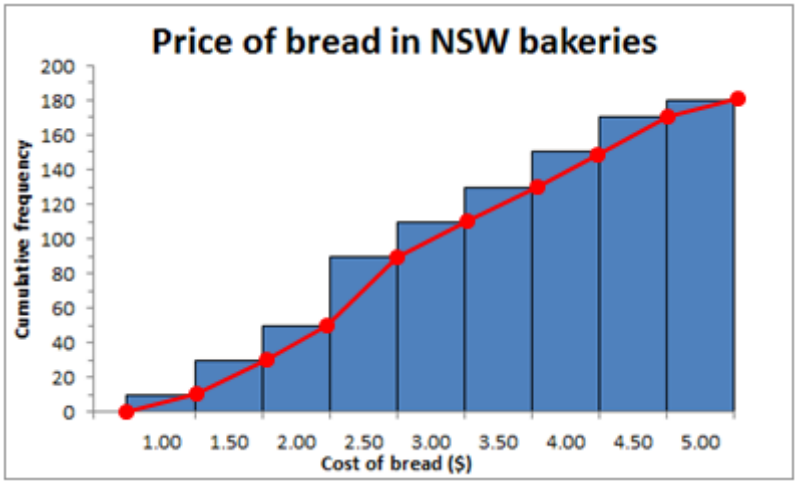
Weight of fish	300	325	350	375	400	425	450	475	500
Frequency	1	4	6	3	5	8	9	3	2

3. A class of students have undergone a statistics test and their results were recorded as follows (out of 10 marks):

10	2	6	7	8	3	5	8	7	1	9	7	8
6	3	8	9	10	6	7	8	7	4	7	3	2

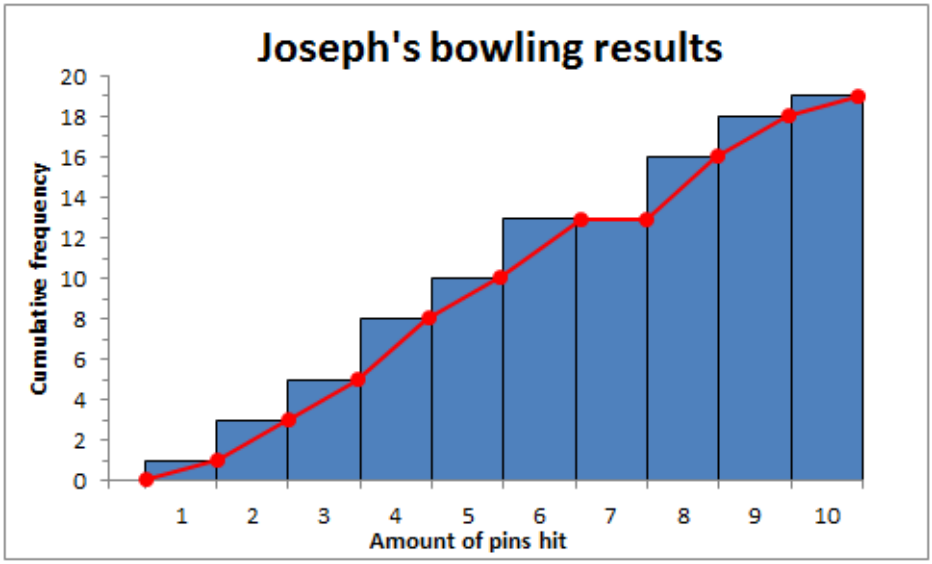
Construct a frequency distribution table and hence, calculate the mean.

4. By using the following ogive, determine the average ($\bar{x}$) price for a loaf of bread in NSW bakeries:



Note: The cumulative frequency is rounded to the nearest 10

5. Joseph went bowling and after a round, his results were shown in the form of an ogive. Determine the average number of pins Joseph hit per bowl by using the ogive as follows:



6. A new "all you can eat" restaurant has a set pricing structure for entry into the buffet room. The pricing is as follows:

Age Group	Under 3	3 - 7	8 - 13	14 - 18	18 +
Entry Fee	Free	$ 5	$ 10	$ 15	$ 20

There was also a sign at the front of the restaurant showing the special opening promotion as follows:

Concession holders will receive a 20% discount for their regular age group entry fee
University students will receive a 10% discount for their regular age group entry fee

By using the following recorded ages from the store:

2	4	17	34	84C	46	37	49*	17	47C	1	38	37
27	36*	98C	37*	26	19	15	14	13	17	18*	18	16
17*	26	28*	16	14	16	17*	47	83C	73	46	13	15
6	3	4	1	3	2	8	43	32	17*	28	15	11C

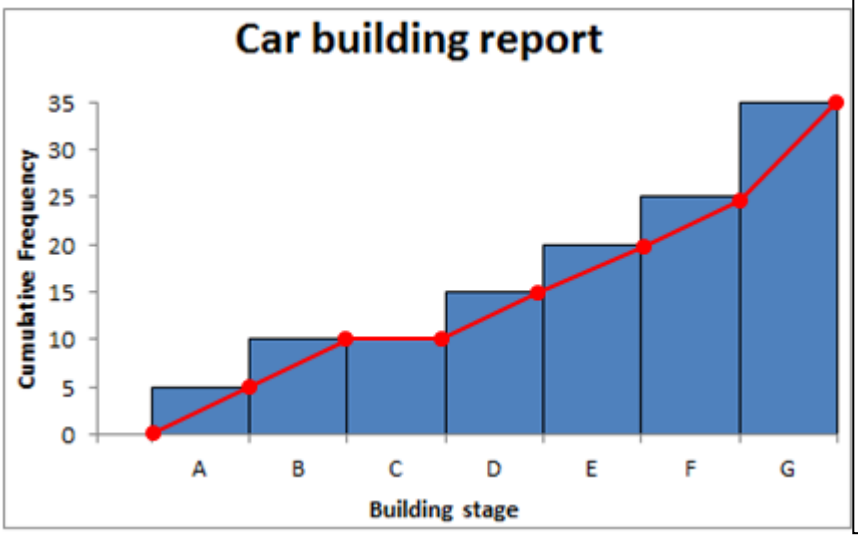
- * denotes University student
- C denotes concession

(i) Construct a frequency distribution table and also record the frequency of students/concessions for that unique age group if applicable

(ii) Hence, calculate the average entry fee the restaurant will receive on opening night (to the nearest dollar)

(iii) Sketch a frequency histogram/polygon and an ogive

7. A manufacturing company was inspecting a batch of cars to see how efficiently the production process was going amongst a group of assemblers. The following ogive was provided to a statistician (all values are rounded to the nearest 5):



Stages
A - 0-19% complete w/o check
B - 20-39% complete w/o check
C - 40-59% complete w/o check
D - 60-79% complete w/o check
E - 80-100% complete w/o check
F - 100% complete but unsatisfactory
G - 100% complete and ready

Ideally, 35 cars should be built per day, but this is quite impossible. To encourage the workers, a promotion will be granted if the statistician finds that the average amount of cars being built per day exceeds 20 per day. The statistician will add bonus points if workers are able to pass more stages during the production of the cars to achieve this promotion but will also deduct points for complete but unsatisfactory cars as follows:

$$\begin{aligned} Number\ of\ cars\ built = {} & Number\ of\ 100\%\ complete\ and\ ready\ cars \\ & +0.5[Number\ of\ stage\ (A+B+C+D+E)\ cars\ w/o\ check] \\ & -Numberof\ 100\%\ complete\ but\ unsatifactory\ cars \end{aligned}$$

Determine if the workers will receive a promotion. Hint: Construct a cumulative frequency distribution table to aid in your calculations.

Grouping Data

1. Determine and explain in the following situations the type of data (discrete or continuous) the recordings belong to:
   (i) Temperature recordings over a period of time
   (ii) The length of strings recorded after they have been cut up a few times
   (iii) The results from counting the number of people in a group
   (iv) The number of heads recorded during a coin toss session
   (v) A collection of the weights of new born babies
   (vi) The results from counting the number of males at a school

2. Fill in the blanks and determine the mean, modal class and median class of the price of leather jackets in a shopping centre.

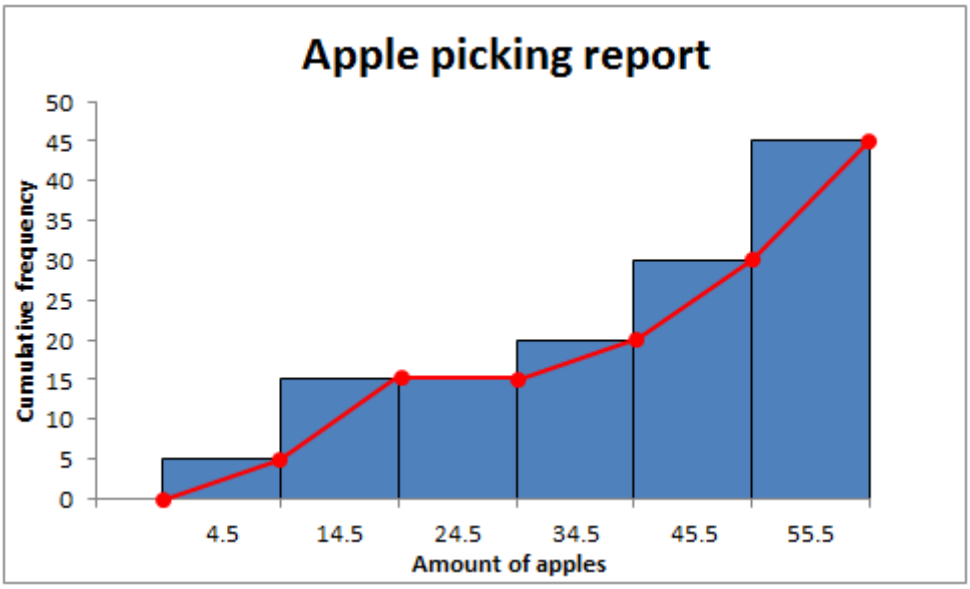
Price ($)	c.c	Tally	f	$f \times c.c.$	$c.f.$
1-50		𝍸			
51-100		𝍸 ||			
101-150		𝍸 𝍸			
151-200		||||			
200-250		|			

3. The following results from a driving stunt test (out of 100) were presented to the head of office to assess the difficulty of the test:

89	48	68	78	96	86	67	57	58	47	38	29	17
47	38	29	17	38	68	98	94	89	79	89	78	78
89	78	64	84	39	87	78	99	58	48	72	87	18

Complete a grouped frequency distribution table with intervals of 20 (beginning with 1-20, 21-40 and so on) including the c.c., tally, f, $f \times c.c.$, $c.f$. Hence, calculate the average score and determine if the test is difficult (an average score of 40% is considered to be difficult).

4. An apple picking report made by a statistician was given to a farmer as he wanted to know how efficient his apple pickers were. By using the report, complete a grouped frequency distribution table including the classes, c.c., f, $f \times c.c.$, $c.f$. Hence, calculate the average amount of apples picked per person and sketch a frequency distribution histogram and polygon.



Note: Cumulative frequency has been rounded to the nearest 5.

5. A corner store owner recorded the number of coke cans sold each day over a period as follows:

5	2	38	26	24	9	6	73	13	17	6	18	23
53	19	1	27	36	29	21	27	39	17	39	47	26

Using this information:

(i) Complete a grouped frequency distribution table with intervals of 5 (beginning with 0-4, 5-9
(ii) and so on) including the c.c., tally, f, $f \times c.c.$, $c.f$.
(iii) Determine the modal class and median class
(iv) How many days did the owner record the results for?
(v) Calculate the average number of cans sold per day.
(vi) Calculate the range.
(vii) Sketch a frequency histogram/polygon and an ogive

6. A school entered a mathematics competition in which A and B graded schools are awarded additional funding from the government. The successful schools have not yet been announced although individual student results were published. The results received are as follows (out of 100):

38	43	84	95	47	39	59	98	18	58	98	74	96
48	49	47	96	57	87	82	78	71	36	82	38	82
37	8	47	83	27	84	92	74	27	75	47	37	92

There are 2 ways in which the school can be eligible for the prize. The school must either be as a whole, an A or B graded school (determined by calculating the average student mark and comparing it to the table below) or have 30%+ of students that have undergone the test been given an A or B grade. The grading system is as follows:

Class of Marks	Individual grade	Overall school Grade
85 - 100	A	A
75 - 84	B	B
65 - 74	C	C
50 - 64	D	D
0 - 49	E	E

The funding is as follows:

- Schools that are A or B graded overall will receive $1000 multiplied by the number of students that have competed
- Schools where 30%+ of students that have competed had received an A or B grade will receive 30% multiplied by $1000 multiplied by the number of students that have competed

Using this information:

(i) Complete a grouped frequency distribution table including the class, c.c., tally, f, $f \times c.c.$, $c.f$.
(ii) Determine the modal class and median class
(iii) Calculate the average mark
(iv) Calculate the range
(v) Sketch a frequency histogram/polygon and an ogive
(vi) Determine if the school qualifies for additional funding. If so, calculate how much the school will receive.'

7. A store owner had a look at his payment booklet as he was curious about the average salary of his employees before and after tax. He had recorded the following salaries before tax (in $ times 1000):

48	26	84	92	72	82	37	28	19	27	93	27	1
37	18	29	84	37	29	17	93	27	6	27	17	71
3	82	4	73	27	19	39	63	85	28	36	53	28

The individual income tax rates are as follows:

Income	Tax rate
\$0 - \$6,000	0%
\$6,001 - \$37,000	10%
\$37,001 - \$80,000	25%
\$80,001 - \$180,000	35%
\$180,001 and over	45%

Using this information:

(i) Complete a grouped frequency distribution table including the class, c.c., tally, f, $f \times c.c.$, $c.f$.
(ii) Determine the modal class and median class
(iii) Calculate the average salary before tax and after tax
(iv) Calculate the range
(v) Sketch a frequency histogram/polygon and an ogive
(vi) What is the average tax rate of his employees? Hint: Find the average salary before tax and compare this with the individual income tax rates table.

End of homework