M1 Maths

S5-1 Statistics from Grouped Data

- calculating statistics from grouped data

Summary Learn Solve Revise Answers

Summary

There are procedures for calculating statistics including modal class, range, mean, median and standard deviation for grouped data by hand.

It is also possible to find these statistics using the statistics functions of a scientific or graphics calculator.

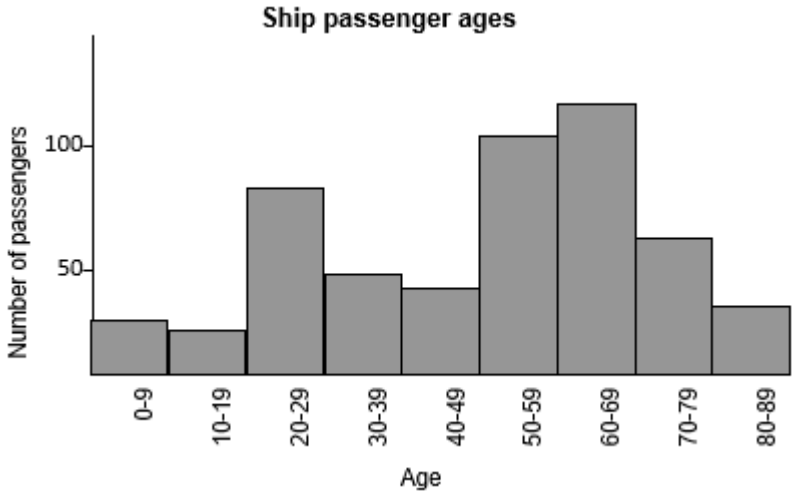
Learn

Grouped data was introduced in Module S3-2. It could be worth going back to that module to refresh your memory. Read the sections on *Grouped Tables* and *Histograms*.

The table below is an example of grouped data. It tells us the ages of the passengers on a ship. Rather than giving us the age of every passenger, it tells us how many are in each age group.

Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
No of passengers	24	21	82	50	48	103	121	68	31

Here is the same data presented as a histogram.



Finding Statistics by Hand

Modal Class

The ranges 0-9, 10-19 etc. are called classes. For grouped data, we talk about a modal class rather than a mode.

To find the modal class, just look for the largest number of passengers in a class on the table or the tallest column on the histogram. For the data above, the modal class is 60-69.

Range

The range is conventionally taken as the difference between the top of the highest populated class and the bottom of the lowest populated class. If we knew all the individual ages, the actual range could be smaller, but with grouped data we cannot tell.

In the example above the range is calculated to be $89 – 0 = 89$.

Mean

To calculate the mean, we assume that all the passengers in a given class are exactly in the middle of that class. So, those in the 0-9 class are all aged 4.5, those in the 40-49 class are all aged 44.5 and so on. We call these the central value for each class.

Then we find the total of the ages for each class by multiplying the central value for that class by the number of people in the class:

$4.5 \times 24 + 14.5 \times 21 + 24.5 \times 82 + 34.5 \times 50 + 44.5 \times 48 + 54.5 \times 103 + 64.5 \times 121 + 74.5 \times 68 + 84.5 \times 31 = 27\,368$

This gives us a number representing the total of the ages of all the passengers.

Then we add the number of passengers in each class to find the total number of passengers:

$24 + 21 + 82 + 50 + 48 + 103 + 121 + 68 + 31 = 548$.

Then we divide the total of the ages by the number of passengers:

$27\,368 \div 548 = 49.97$

As the actual average age could be a bit different from this, we round the result reasonably. 50 would be a good answer.

Median

To find the median, first we determine the number of passengers. We have already

seen that this is 548. The median is then half way between passenger 274 and passenger 275, i.e. 274.5.

Then we count passengers in each class until we get to just before and just after 274.5.

24 + 21 + 82 + 50 + 48 = 225

24 + 21 + 82 + 50 + 48 + 103 = 328

So 274.5 lies in the 50-59 class. We then find the fraction of the way through the class. This is given by

(274.5 – 225) ÷ (328 – 225) = 0.48 of the way.

Now, as ages are continuous, we take the class boundaries as 49.5 and 59.5. The width of the class is then 10. We find 0.48 of 10 which is 4.8. Then we add this to the lower boundary: 49.5 + 4.8 = 54.3.

As ages are given as whole numbers, we round this. So the median age is taken as 54.

Standard deviation

To work out the standard deviation, we would, again, assume that all ages were in the centre of their class and then calculate the standard deviation of the 548 ages. As many of the ages would be the same, this is not as onerous as it first appears.

We have already calculated the mean to be 50 and the total number of passengers to be 548. We then proceed as shown below. [Note the row labelled 'Sum' is the sum of the squared deviations for that class, i.e. the squared deviation multiplied by the number of passengers.]

Age class	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
No of passengers	24	21	82	50	48	103	121	68	31
Central value	4.5	14.5	24.5	34.5	44.5	54.5	64.5	74.5	84.5
Deviation	−45.5	−35.5	−25.5	−15.5	−5.5	4.5	14.5	24.5	34.5
Squared deviation	2070.25	1260.25	650.25	240.25	30.25	20.25	210.25	600.25	1190.25
Sum	49686	26465	53321	12013	1452	2086	25440	40817	36898

Total squared deviation (the sum of the bottom row) = 248 178

Mean squared deviation = 248 178 ÷ 548 = 452.9

Root mean square deviation = $\sqrt{452.9}$ = 21.3

So the standard deviation is 21.3.

Again, as the individual ages are not known, this will be an approximation to the actual standard deviation, do we might round it to 21.

Practice

Q1 Hourly pay in dollars for part-time jobs of students at St Cuthbert's High School:

Pay	0-5	5-10	10-15	15-20	20-25	25-30
No of students	1	12	65	46	14	3

For the data above, find the following by hand:

(a) the modal class
(b) the range
(c) the mean
(d) the median
(e) the standard deviation

Finding Statistics with a Calculator

Finding statistics on grouped data using the statistics functions of a calculator is quite straightforward and you should find out how to enter grouped data on your calculator and how to generate statistics from it.

There are a few things you will need to do by hand including

- Finding the modal class
- Finding the range
- Finding the central value for each class.

Practice

Q2 For the data in Question P1, use a calculator to find:
(a) the mean
(b) the median
(c) the standard deviation

Solve

Q51 In the table for the ages of boat passengers above, the classes were 0-9, 10-19 etc. where the top of one class is different from the bottom of the next class.

In the table for the hourly pays for part time jobs above, the classes were 0-5, 5-10 etc. where the top of one class is the same as the bottom of the next class.

Explain why the two tables would have been constructed differently.

Revise

Revision Set 1

Q61 The following table shows the number of visits to the pub in the past month by the residents of Codgers Retirement Home.

Visits	0-4	5-9	10-14	15-19	20-24	25-29
No of residents	21	13	6	7	3	9

(a) Find the modal class, range, mean, median and standard deviation by hand.

(b) Find the mean, median and standard deviation using your calculator.

Answers

Q1 (a) 10-15 (b) 30 (c) \$14.95 (d) \$14.46
(e) \$1.76

Q2 (a) \$14.95 (b) \$14.46 (c) \$1.76

Q51 Age, the way it is usually quoted, is a discrete variable. Wage, on the other hand is a continuous variable. It may be argued that money is discrete because you can't give someone half a cent in cash. But rates per hour can be quoted to a fraction of a cent, then the final wage for the fortnight rounded to the nearest cent.

Q61 (a) modal class 5-9; range 29; mean 10.7; median 7; standard deviation 9.05.
(c) mean 10.7; median 7; standard deviation 9.05.