Grouped Data

What is grouped data?

- Grouped data is a way of organising raw data into categories or intervals to make it easier to analyse and interpret

- Instead of listing individual values, the data is grouped into classes or ranges, often displayed in a frequency table.

Grade %	Frequency
0 – 9	3
10 – 19	4
20 – 29	2
30 – 39	0
40 – 49	6
50 – 59	4
60 – 69	7
70 – 79	2
80 – 89	1
90 – 100	3

Understanding a frequency table

Grade %	Frequency
0 – 9	3
10 – 19	4
20 – 29	2
30 – 39	0
40 – 49	6
50 – 59	4
60 – 69	7
70 – 79	2
80 – 89	1
90 – 100	3

These are the groups

These are the frequencies (the number of things in this group)

Calculating the Mean

- We cannot work out an exact mean for grouped data because we don't know the individual data values in the dataset
- We can however estimate the mean
- We know that the mean of a group of data is:
  - $\frac{Sum}{No.\ of\ items}$

Grade %	Frequency
0 – 9	3
10 – 19	4
20 – 29	2
30 – 39	0
40 – 49	6
50 – 59	4
60 – 69	7
70 – 79	2
80 – 89	1
90 – 100	3

Calculating the Mean

- For grouped data we calculate the mean using midpoints of groups
- We use this equation for the mean of grouped data:
  - $\frac{\sum[f*m]}{\sum f}$
  - Where f is the frequencies and m is the midpoints

Grade %	Frequency
0 – 9	3
10 – 19	4
20 – 29	2
30 – 39	0
40 – 49	6
50 – 59	4
60 – 69	7
70 – 79	2
80 – 89	1
90 – 100	3

Calculating midpoints

- We can work out the midpoint of a group of data by taking the maximum value, adding the minimum value then dividing by 2
- $m = \frac{maximum+minimum}{2}$
- When working with a frequency table we put this in a third column to make it easier to track

Grade %	Frequency	Midpoint
0 – 9	3	4.5
10 – 19	4	14.5
20 – 29	2	24.5
30 – 39	0	34.5
40 – 49	6	44.5
50 – 59	4	54.5
60 – 69	7	64.5
70 – 79	2	74.5
80 – 89	1	84.5
90 – 100	3	95

Calculating the Mean

- The next step towards calculating our mean is to total up our frequency
- For our example we can write this out as:
  - $\sum f = 3 + 4 + 2 + 0 + 6 + 4 + 7 + 2 + 1 + 3$
- We can put this at the bottom of the table
- We can also think about what this means in the scenario, so in this example there must be 32 students in the class.

Grade %	Frequency	Midpoint
0 – 9	3	4.5
10 – 19	4	14.5
20 – 29	2	24.5
30 – 39	0	34.5
40 – 49	6	44.5
50 – 59	4	54.5
60 – 69	7	64.5
70 – 79	2	74.5
80 – 89	1	84.5
90 – 100	3	95
	32	

Calculating the Mean

- Next, we need to work out the top part of our fraction:
  - $\sum[f * m]$
- It's best to again add another column to our table to keep track of our values

Grade %	Frequency	Midpoint	F*M
0 – 9	3	4.5	13.5
10 – 19	4	14.5	58
20 – 29	2	24.5	49
30 – 39	0	34.5	0
40 – 49	6	44.5	267
50 – 59	4	54.5	218
60 – 69	7	64.5	451.5
70 – 79	2	74.5	149
80 – 89	1	84.5	84.5
90 – 100	3	95	285
	32		

Calculating the Mean

- Finally, we can work out the sum of our F*M column ($\sum[f * m]$)
- $\sum[f * m] =$ 13.5 + 58 + 49 + 0 + 267 + 218 + 451.5 + 149 + 84.5 + 285
- $\sum[f * m] = 1575.5$
- Finally, we can put this at the bottom of our table

Grade %	Frequency	Midpoint	F*M
0 – 9	3	4.5	13.5
10 – 19	4	14.5	58
20 – 29	2	24.5	49
30 – 39	0	34.5	0
40 – 49	6	44.5	267
50 – 59	4	54.5	218
60 – 69	7	64.5	451.5
70 – 79	2	74.5	149
80 – 89	1	84.5	84.5
90 – 100	3	95	285
	32		1575.5

Calculating the Mean

- Now we have both the top and bottom of our fraction we can calculate the mean

- Mean = $\frac{\sum[f*m]}{\sum f} = \frac{1575.5}{32} = 49.23$

Grade %	Frequency	Midpoint	F*M
0 – 9	3	4.5	13.5
10 – 19	4	14.5	58
20 – 29	2	24.5	49
30 – 39	0	34.5	0
40 – 49	6	44.5	267
50 – 59	4	54.5	218
60 – 69	7	64.5	451.5
70 – 79	2	74.5	149
80 – 89	1	84.5	84.5
90 – 100	3	94.5	285
	32		1575.5

Finding the mode

- Finding a mode is super easy, we just look for the highest frequency
- So, for our example our highest frequency is 7 which is for the group 60-69%
- So, the mode must exist between 60% and 69%

Grade %	Frequency	Midpoint	F*M
0 – 9	3	4.5	13.5
10 – 19	4	14.5	58
20 – 29	2	24.5	49
30 – 39	0	34.5	0
40 – 49	6	44.5	267
50 – 59	4	54.5	218
60 – 69	7	64.5	451.5
70 – 79	2	74.5	149
80 – 89	1	84.5	84.5
90 – 100	3	94.5	285
	32		1575.5

Finding the median

- The first step in finding the median is to expand our table again to have a cumulative frequency column
- The cumulative frequency is a running total of frequencies

Grade %	Frequency	Midpoint	F*M	cf
0 – 9	3	4.5	13.5	3
10 – 19	4	14.5	58	7
20 – 29	2	24.5	49	9
30 – 39	0	34.5	0	9
40 – 49	6	44.5	267	15
50 – 59	4	54.5	218	19
60 – 69	7	64.5	451.5	26
70 – 79	2	74.5	149	28
80 – 89	1	84.5	84.5	29
90 – 100	3	94.5	285	32
	32		1575.5	

Finding the median

- Finally, we do $\frac{1}{2}(\sum f)$ which is $\frac{1}{2}(32)$
- So we know the median value must be the 16th value
- We know 16 is between 15 and 19
- As the row 50-59 covers the values between 15 and 19 then that must be our median
- So, our median is in the range 50-59

Grade %	Frequency	Midpoint	F*M	cf
0 – 9	3	4.5	13.5	3
10 – 19	4	14.5	58	7
20 – 29	2	24.5	49	9
30 – 39	0	34.5	0	9
40 – 49	6	44.5	267	15
50 – 59	4	54.5	218	19
60 – 69	7	64.5	451.5	26
70 – 79	2	74.5	149	28
80 – 89	1	84.5	84.5	29
90 – 100	3	94.5	285	32
	32		1575.5	

Your Turn

- Can you work out the mean, median and mode for this grouped data:
- This group is counting the weight of items produced from a manufacturing line in an hour

Weight (g)	Frequency	Midpoint	F*M	cf
0 – 49	46			
50 – 99	72			
100 – 149	149			
150 – 199	641			
200 – 249	527			
250 – 299	266			
300 – 349	19			
350 – 400	2			

Your Turn - Answers

- Mean = $\frac{324160}{1722} = 188.246\text{g}$
- Median = $\frac{1722}{2} = 861$ (150 - 199)g
- Mode = (150 – 199)g

Weight (g)	Frequency	Midpoint	F*M	Cf
0 – 49	46	24.5	1127	46
50 – 99	72	74.5	5364	118
100 – 149	149	124.5	18550.5	267
150 – 199	641	174.5	111854.5	908
200 – 249	527	224.5	118311.5	1435
250 – 299	266	274.5	62037	1701
300 – 349	19	324.5	6165.5	1720
350 – 400	2	375	750	1722
	1722		324160	

Standard Deviation

- Standard Deviation is a method used to measure how spread-out data is
- It measures how far data points are from the average (mean)
- The standard formula is:
- $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
- Where:
- x = the data values
- $\bar{x}$ = the mean
- n = the number of items

Calculating Standard Deviation

- The first step towards finding the standard deviation of grouped data is to rearrange the formula for standard deviation

- $s = \sqrt{\frac{\sum f(m-\bar{x})^2}{n-1}}$

- Where:
  - m = the midpoint of a range
  - $\bar{x}$ = the mean
  - n = the number of items
  - f = the frequency of that range

Calculating Standard Deviation

- The first step towards finding the standard deviation is to find the mean
- Which starts with finding the midpoint of our data
- We again do this by doing:
  - $\frac{maximum + minimum}{2}$

Length	Frequency	Midpoint
20 – 24	45	22
25 – 29	53	27
30 – 34	99	32
35 – 39	256	37
40 – 44	621	42
45 – 49	34	47
50 – 54	21	52
55 – 60	2	57.5

Calculating Standard Deviation

- Now we have the midpoints do the following things to work out the mean:
  - Put in our f*m column
  - Put in our totals row at the bottom
  - Work out the mean
- Mean = $\frac{43948}{1131}$ = 38.8576

Length	Frequency	Midpoint	f*m
20 – 24	45	22	990
25 – 29	53	27	1431
30 – 34	99	32	3168
35 – 39	256	37	9472
40 – 44	621	42	26082
45 – 49	34	47	1598
50 – 54	21	52	1092
55 – 60	2	57.5	115
	1131		43948

Calculating Standard Deviation

- Again, to make the process easier on ourselves we should put the mean into our table
- We can then add a column with $m - \bar{x}$ which we know is important for our equation

Length	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$
20 – 24	45	22	990	38.8576	-16.8576
25 – 29	53	27	1431	38.8576	-11.8576
30 – 34	99	32	3168	38.8576	-6.8576
35 – 39	256	37	9472	38.8576	-1.8576
40 – 44	621	42	26082	38.8576	3.1424
45 – 49	34	47	1598	38.8576	8.1424
50 – 54	21	52	1092	38.8576	13.1424
55 – 60	2	57.5	115	38.8576	18.6424
	1131		43948		

Calculating Standard Deviation

- Next we can add a column squaring the $m - \bar{x}$ value, getting us closer to the top of the equation

Length	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$	$(m - \bar{x})^2$
20 – 24	45	22	990	38.8576	-16.8576	284.1787
25 – 29	53	27	1431	38.8576	-11.8576	140.6027
30 – 34	99	32	3168	38.8576	-6.8576	47.0267
35 – 39	256	37	9472	38.8576	-1.8576	3.4507
40 – 44	621	42	26082	38.8576	3.1424	9.8747
45 – 49	34	47	1598	38.8576	8.1424	66.2987
50 – 54	21	52	1092	38.8576	13.1424	172.7227
55 – 60	2	57.5	115	38.8576	18.6424	347.5391
	1131		43948			

Calculating Standard Deviation

- We can add yet another column to work out $f(m-\bar{x})^2$

Length	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$	$(m-\bar{x})^2$	$f(m-\bar{x})^2$
20 – 24	45	22	990	38.8576	-16.8576	284.1787	12788.0415
25 – 29	53	27	1431	38.8576	-11.8576	140.6027	7451.9431
30 – 34	99	32	3168	38.8576	-6.8576	47.0267	4655.6433
35 – 39	256	37	9472	38.8576	-1.8576	3.4507	883.3792
40 – 44	621	42	26082	38.8576	3.1424	9.8747	6132.1887
45 – 49	34	47	1598	38.8576	8.1424	66.2987	2254.1558
50 – 54	21	52	1092	38.8576	13.1424	172.7227	3627.1767
55 – 60	2	57.5	115	38.8576	18.6424	347.5391	695.0782
	1131		43948				

Calculating Standard Deviation

- For the top of our equation, we need $\sum f(m-\bar{x})^2$ so we should put the total in the total row

Length	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$	$(m-\bar{x})^2$	$f(m-\bar{x})^2$
20 – 24	45	22	990	38.8576	-16.8576	284.1787	12788.0415
25 – 29	53	27	1431	38.8576	-11.8576	140.6027	7451.9431
30 – 34	99	32	3168	38.8576	-6.8576	47.0267	4655.6433
35 – 39	256	37	9472	38.8576	-1.8576	3.4507	883.3792
40 – 44	621	42	26082	38.8576	3.1424	9.8747	6132.1887
45 – 49	34	47	1598	38.8576	8.1424	66.2987	2254.1558
50 – 54	21	52	1092	38.8576	13.1424	172.7227	3627.1767
55 – 60	2	57.5	115	38.8576	18.6424	347.5391	695.0782
	1131		43948				38487.6065

Calculating Standard Deviation

- Now we have our values we can put them into the equation from earlier:
- $s = \sqrt{\frac{\sum f(m-\bar{x})^2}{n-1}}$
- We know the top is 38487.6065 and the bottom is 1131 – 1
- So we can write out:
- $s = \sqrt{\frac{38487.6065}{1131-1}} = 5.8361$

Length	Frequency	Midpoint	f*m	$f(m-\bar{x})^2$
20 – 24	45	22	990	12788.0415
25 – 29	53	27	1431	7451.9431
30 – 34	99	32	3168	4655.6433
35 – 39	256	37	9472	883.3792
40 – 44	621	42	26082	6132.1887
45 – 49	34	47	1598	2254.1558
50 – 54	21	52	1092	3627.1767
55 – 60	2	57.5	115	695.0782
	1131		43948	38487.6065

Your Turn

- Can you work out the standard deviation for this dataset.

Price (£)	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$	$(m-\bar{x})^2$	$f(m-\bar{x})^2$
0 – 0.19	6						
0.20 – 0.39	211						
0.40 – 0.59	321						
0.60 – 0.79	64						
0.80 – 0.99	9						
1.00 – 1.19	4						
1.20 – 1.39	0						
1.40 – 1.59	1						

- This dataset is recording the price of a component from multiple different sources

Your Turn - Answers

Price (£)	Frequency	Midpoint	f*m	$\bar{x}$	m-$\bar{x}$	$(m-\bar{x})^2$	$f(m-\bar{x})^2$
0 – 0.19	6	0.095	0.57	0.4547	-0.3597	0.1294	0.7764
0.20 – 0.39	211	0.295	62.245	0.4547	-0.1597	0.0255	5.3805
0.40 – 0.59	321	0.495	158.895	0.4547	0.0403	$1.6240 * 10^{-3}$	0.5213
0.60 – 0.79	64	0.695	44.48	0.4547	0.2403	0.0577	3.6928
0.80 – 0.99	9	0.895	8.055	0.4547	0.4403	0.1939	1.7451
1.00 – 1.19	4	1.095	4.38	0.4547	0.6403	0.4100	1.64
1.20 – 1.39	0	1.295	0	0.4547	0.8403	0.7061	0
1.40 – 1.59	1	1.495	1.495	0.4547	1.0403	1.0822	1.0822
	616		280.12	0.4547			14.1783

Standard deviation = $\sqrt{\frac{14.1783}{616-1}} = 0.1518359191 \approx 0.1518$