Measures of Central Tendency and Dispersion for Grouped Data

Grouped Data

In the previous sections all data were given in a *data array,* i.e. they were listed in a row, in a column, or they were listed in a table where each cell contained one observation. When we have large number of observations we can group the data in a two column table, where *in the first column we list all possible values,* and in the second column we list the frequencies at which the different values occur.

For example let's consider the following 23 scores:
12, 10, 8, 9, 14, 13, 12, 10, 8, 8, 8, 8, 9, 14, 13, 13, 14, 14, 14, 9, 10, 16, 6

Some values occur once, some values twice, some values several times. First we create a column for counting (tally), and for the frequencies:

Values	**Tally**	**Frequency**
6	/	1
8	/////	5
9	///	3
10	///	3
12	//	2
13	///	3
14	/////	5
16	/	1

Next, we omit the Tally column, so our final table looks like this:

Values (x)	**Frequency (f)**
6	1
8	5
9	3
10	3
12	2
13	3
14	5
16	1

The same information can be further summarized in a *class frequency table,* where instead of the exact values we define non overlapping intervals for the values. For example

Interval (From *Low to High*)	Frequency (f)
From 0 to less than 5	0
From 5 to less than 10	9
From 10 to less than 15	13
From 15 to less than 20	1

Each interval is characterized by its *middle point,* so if we have a class frequency table we identify the middle points of the intervals and use those in all calculations as values.

To get the middle point (mid value) of the intervals we use the following formula:

$$\text{midpoint} = \frac{Lower\ limit + Higher\ limit}{2}$$

So the final table:

Interval	Midpoints (x)	Frequency (f)
From 0 to less than 5	$\frac{0+5}{2} = 2.5$	0
From 5 to less than 10	$\frac{5+10}{2} = 7.5$	9
From 10 to less than 15	$\frac{10+15}{2} = 12.5$	13
From 15 to less than 20	$\frac{15+20}{2} = 17.5$	1

Important: When we have *class frequency distributions,* the *class middle points become the values* we use: ***x* = midpoint**

The Arithmetic Mean

Grouped data refer to frequency distributions. Three cases are covered in this category. The first case deals with classes consisting of one observation (discrete series, see example 2). The second has class intervals consisting of a range of values, which may be discrete or continuous (examples 3 and 4). Just a reminder, a discrete variable is the result of a count, whereas a continuous variable is the result of a measurement. The formulas to apply are very similar to the weighted mean with slight adjustments.

$$\textbf{\textit{arithmetic mean}} = \mu = \frac{\sum f x}{N}$$

Example 1

A storeowner took an inventory count and found that he had sold 124 dresses during the Eid season. The numbers of dresses sold according to dress size were as follows:

Size	Number of Dresses
4	3
5	10
6	18
7	20
8	24
9	20
10	14
11	10
12	5
Total	124

Determine the mean size of the dresses sold.

Solution:

$$\boldsymbol{\mu} = \frac{\sum fx}{N} = \frac{4x3+5x10+6x18+7x20+8x24+9x20+10x14+11x10+12x5}{124} = \frac{992}{124} = \mathbf{8}$$

Example 2

The following table contains a frequency distribution of ages from a sample of Emarat FM radio station listeners. What is the mean age of an Emarat FM listener based on these data?

Age	Frequency
10 and under 15	6
15 and under 20	22
20 and under 25	35
25 and under 30	29
30 and under 35	16
35 and under 40	8
40 and under 45	4
45 and under 50	2
Total	122

Solution:

Age	Midpoint (x)	Frequency
$10 \leq$ age < 15	12.5	6
$15 \leq$ age < 20	17.5	22
$20 \leq$ age < 25	22.5	35
$25 \leq$ age < 30	27.5	29
$30 \leq$ age < 35	32.5	16
$35 \leq$ age < 40	37.5	8
$40 \leq$ age < 45	42.5	4
$45 \leq$ age < 50	47.5	2
Total		122

$$\boldsymbol{\mu} = \frac{\sum fx}{N} = \frac{6x12.5+22x17.5+35x22.5+29x27.5+16x32.5+8x37.5+4x42.5+2x47.5}{122} = \frac{3130}{122} = \mathbf{25.66}$$

The mean age of this FM listener is 22.66 years.

In the latter case the answer for the mean is an approximate answer. We are making the assumption that the mean of the values in a class interval equals the class midpoint. In some cases, the values in a class interval may actually fall at one end of the interval; in those cases, the grouped mean contains considerable error.

Example 3

Find the mean of the following data that represent the number of telephone appointments made per fifteen-minute intervals.

Number of Appointments	Frequency
0-1	31
2-3	57
4-5	26
6-7	14
8-9	6
10-11	3

Solution:

This is an example of a *discrete series* (result of a count) where the classes in the distribution consist of a range of values (each class consists of 2 values).

Number of Appointments	Middle points *x*	Frequency *f*
0-1	0.5	31
2-3	2.5	57
4-5	4.5	26
6-7	6.5	14
8-9	8.5	6
10-11	10.5	3
	Total	**137**

$$\mu = \frac{\sum fx}{N} = \frac{448.5}{137} = 3.2727$$

The mean number of telephone appointments made per fifteen-minute intervals is 3.2727.

Example 4

Suppose that we wish to find the mean age of a sample of ten students whose ages are as follows: 19, 19, 20, 20, 20, 21, 21, 21, 21, and 22.

Solution:

There are two 19s, three 20s, four 21s, and one 22. Multiplying each different age by its frequency and adding the products yields:

$$\mu = \frac{\sum fx}{N} = \frac{2x19+3x20+4x21+22}{2+3+4+1} = \frac{204}{10} = 20.4$$

The Weighted Mean

In the preceding definition of the mean, every score is treated equally, but there are cases in which the scores vary in their degree of importance. In such cases, we can calculate the mean by applying different **weights** to different scores. A weight is a value corresponding to how much the score is weighed. Given a list of scores (observations) x_1, x_2, x_3,...,x_n and a corresponding list of weights w_1, w_2, w_3,..., w_n , the **weighted mean** is obtained by using the formula:

$$\textit{weighted mean} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n}{w_1 + w_2 + w_3 + \cdots + w_n} = \frac{\sum wx}{\sum w}$$

For *class frequency tables* we have to use the middle points for the *x* values, otherwise the formula is the same.

Example 5

A weighted mean is frequently used in the determination of a final average for a course that may include 4 tests plus a final examination. If the test grades are 55, 80, 63, 82, and the exam mark is 90, the arithmetic mean of 74 does not reflect the greater importance placed on the final exam. Suppose the instructor allocates the following weights to these tests: 10%, 20%, 15%, 15% respectively, and 40% to the final exam. Find the weighted mean.

Solution:

Using the given scores and weights, we calculate the weighted mean by applying the above formula.

$$\textit{weighted mean} = \frac{\sum wx}{\sum w} = \frac{55x10 + 80x20 + 63x15 + 82x15 + 90x40}{10+20+15+15+40} = \frac{7925}{100} = 79.25$$

The Mode

The second measure of central tendency that we can get for grouped data is the mode. To find the most frequent value we simply have to choose the *x* or the *m* value which occurs with the greatest frequency. *No calculation is needed.*

Example 6

Identify the *Modes* in the first 4 Examples.

Solution:

Example	Highest frequency	Mode
1	24	Size 8
2	35	22.5 years
3	57	2.5 appointments
4	4	21 years

The Median

The median is the middle item when the data are arranged in order of size (e.g. from least to greatest).

If there is an ***even*** number of items in the data (*N*), then the median is the average of the two middle values.
If *N* is an ***odd*** number we get the median immediately.

Example 7

Identify the *Medians* in the first 4 Examples**.**

Solution:

Example	Total frequency	Total frequency / 2	Place of median	Median
1	124	62	62^{nd} and 63^{rd}	8
2	122	61	61^{st} and 62^{nd}	22.5
3	137	68.5	69^{th}	57
4	10	5	5^{th} and 6^{th}	20.5

The Range

The Range is the difference between the smallest and the largest value among the data.

Range = Maximum - Minimum

The Variance

The **variance for grouped data** or data arranged in the form of a frequency distribution is given by

$$Variance = \sigma^2 = \frac{\sum f(x - \mu)^2}{N}$$

where ***x*** is the observation in the class interval containing a single value, and the midpoint of the class interval consisting of a range of values,
μ is the arithmetic mean,
f is the frequency in the class interval,
and ***N*** is the total number of observations.

Example 8 – Discrete Case

Calculate the variance for the following table. Round the final answer to 2dp.

Size *x*	**Frequency** *f*
7	5
8	8
9	3
10	12
11	14

Solution: Add all the frequencies: N = 42

$$\mu = \frac{\sum fx}{N} = \frac{5x7+8x8+3x9+12x10+14x11}{42} = \frac{400}{42} = \frac{200}{21}$$

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{5\left(7-\frac{200}{21}\right)^2+8\left(8-\frac{200}{21}\right)^2+3\left(9-\frac{200}{21}\right)^2+12\left(10-\frac{200}{21}\right)^2+14\left(11-\frac{200}{21}\right)^2}{42} = \mathbf{2.01}$$

Example 9 – Continuous Case

Compute the variance of the following hourly wages of 24 employees.

Hourly Wages (Dh)	**Number of employees**
20 and under 25	3
25 and under 30	5
30 and under 35	9
35 and under 40	4
40 and under 45	3

Solution:

Class Interval	**Midpoint** *x*	*f*
20 and under 25	22.5	3
25 and under 30	27.5	5
30 and under 35	32.5	9
35 and under 40	37.5	4
40 and under 45	42.5	3
Total		24

$$\mu = \frac{\sum fx}{N} = \frac{3x22.5+5x27.5+9x32.5+4x37.5+3x42.5}{24} = \frac{775}{24}$$

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{3\left(22.5-\frac{775}{24}\right)^2+5\left(27.5-\frac{775}{24}\right)^2+9\left(32.5-\frac{775}{24}\right)^2+4\left(37.5-\frac{775}{24}\right)^2+3\left(42.5-\frac{775}{24}\right)^2+}{24} = \mathbf{34.33}$$

The Standard Deviation

The standard deviation is defined as the square root of the variance. If the variance was calculated in a previous step, its square root provides the answer for the standard deviation. By the same token, if the standard deviation was calculated in a previous step, its squared value gives the answer for the variance.

The **standard deviation for grouped data** or data arranged in the form of a frequency distribution is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma f(x-\mu)^2}{N}}$$

where x is the observation in the class interval containing a single value, and the midpoint of the class interval consisting of a range of values,

μ is the mean,

f is the frequency in the class interval, and

N is the total number of observations.

When we have class frequency distribution we have to use the midpoints and use them as the *x* values, otherwise the formula is the same:

Example 10

Calculate the standard deviation for the following table. Give your answer with 2dp.

Size *x*	**Frequency** *f*
7	5
8	8
9	3
10	12
11	14

Solution: Add all the frequencies: N = 42

$$\mu = \frac{\sum fx}{N} = \frac{5x7+8x8+3x9+12x10+14x11}{42} = \frac{400}{42} = \frac{200}{21}$$

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{5(7-\frac{200}{21})^2 + 8(8-\frac{200}{21})^2 + 3(9-\frac{200}{21})^2 + 12(10-\frac{200}{21})^2 + 14(11-\frac{200}{21})^2}{42} = 2.011337869$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.011337869} = 1.418216 = 1.42$$

Example 11

Compute the standard deviation of the hourly wages of the 24 employees who work at A-Max.

Hourly Wages (Dh)	Number of employees
20 and under 25	3
25 and under 30	5
30 and under 35	9
35 and under 40	4
40 and under 45	3

Solution:

Class Interval	Midpoint x	f
20 and under 25	22.5	3
25 and under 30	27.5	5
30 and under 35	32.5	9
35 and under 40	37.5	4
40 and under 45	42.5	3
Total		24

$$\boldsymbol{\mu} = \frac{\sum f x}{N} = \frac{3x22.5+5x27.5+9x32.5+4x37.5+3x42.5}{24} = \frac{775}{24}$$

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{3(22.5-\frac{775}{24})^2+5(27.5-\frac{775}{24})^2+9(32.5-\frac{775}{24})^2+4(37.5-\frac{775}{24})^2+3(42.5-\frac{775}{24})^2+}{24} = 34.331597$$

From here the *standard deviation* is:

$$\sigma = \sqrt{\sigma^2} = \sqrt{34.331597} = 5.86$$

Exercises 2.3

1. Compute the mean, mode, range and standard deviation for the following continuous variable:

Class intervals	Frequency
From 0 to less than 2	39
From 2 to less than 4	27
From 4 to less than 6	16
From 6 to less than 8	15
From 8 to less than 10	10
From 10 to less than 12	8
From 12 to less than 14	6

2. Two Math classes took the same test, The first class had 18 students. Its mean was 70.5. The second class had 33 students and the class mean was 64. Calculate the weighted mean of the test, based on the results of all 51 students.

3. The mean of 100 students' grades was found to be 40. Later, it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score.

4. State whether the following statements are true or false. Explain your reasoning.

   (a) If one section of 40 students gets an average mean of 60 marks and another section of 60 students obtains a mean of 40 marks, the average mean mark of 100 students is 50.

   (b) The mean is always an accurate indicator of the center of a set of data.

5. Insurance companies continually research ages at death and causes of death. The given frequency table summarizes ages at death for all people who died from gunfire in America during the week of May 15, 1992 period (based on a Time magazine study).

Age at Death	Frequency
From 16 to less than 26	22
From 26 to less than 36	10
From 36 to less than 46	6
From 46 to less than 56	2
From 56 to less than 66	4
From 66 to less than 76	5
From 76 to less than 86	1

Compute the mean, the mode, the range and the standard deviation

6. In a small bank in Boston, the distribution of deposit balances and the number of deposits in a month are as follows:

Deposit Balance (Dollars)	**Number of Deposits**
0 up to 100	25
100 up to 250	100
250 up to 400	175
400 up to 500	74
500 up to 550	66
550 up to 600	35
600 up to 800	5
800 up to 900	18
900 up to 1000	2
Total	500

Calculate the mean and the mode deposit.

7. Compute the range and the mean for the following distribution, which gives the annual profits of 20 small business firms.

Profits (in thousands of dollars)	**Number of firms**
60 and under 70	2
70 and under 80	5
80 and under 90	9
90 and under 100	3
100 and under 110	1

8. Telephone orders received by a catalogue shopping company during the first 4 hours of a business day are as follows:

	1^{st} hour	2^{nd} hour	3^{rd} hour	4^{th} hour
Number of orders	3	5	7	9

Compute the variance.

9. Find the mean and standard deviation for the following discrete variable:

Size	3-4	5-6	7-8	9-10	11-12	13-14	15-16
Frequency	3	7	22	60	85	32	8

10. Calculate the standard deviation and the mean for the following table:

Size	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

11. The following data give information on the number of "no shows" on the daily 6:00 p.m. flight from New York to Boston for records kept for 60 days.

Number of no shows	**Number of days**
0	33
1	14
2	6
3	4
4	2
5	1

Compute the mean, the mode, the range and the standard deviation for the above data to 2 decimal places.

12. A survey is carried out on the number of cars owned by different families. The following results are obtained:

Number of cars, x	0	1	2	3	4
Number of families, f	5	12	14	6	3

Calculate the mean and the standard deviation of the number of cars per family.

13. An investigation of the number of sales made in a month by the sales force of a company is made:

Number of sales, x	0-4	5-9	10-14	15-19	20-24
Number of salespeople, f	1	5	9	12	3

Find the mean and standard deviation of the number of sales made per person.

14. What can be said about a set of numbers whose mean is 95 and whose standard deviation is 0?

15. The following three sets of numbers each have a mean of 50. Which has the largest standard deviation? Which has the smallest standard deviation? (Do not perform any calculations!)

Set X:	42	46	50	54	58
Set Y:	40	45	50	55	60
Set Z:	30	40	50	60	70

16. What would happen to the standard deviation and variance of a set of data if each value was increased by 4?

17. 50 apples taken from an apple tree were measured for size. The results are:

Diameter in cm	8.8	8.9	9.0	9.1	9.2
Number of Apples	7	15	20	5	3

Compute the mean, variance and standard deviation of the diameter of the 50 apples.

Answers: Exercise 2.3

1 mean = 4.64 to 2 d.p. standard deviation = 3.65 to 2 d.p.
mode = 1 range = 12

2 mean = 66.29

3 corrected mean = 39.7

4 (a) false, average = 48
(b) false, not in the presence of extreme values

5 mean = 36 standard deviation = 18.138 mode = 21
range = 60

6 mean = 368.8 mode = 325

7 mean = Dh 83 000 range = Dh 40 000

8 var. = 1.08 to 2 d.p. range = 3

9 mean = 10.68 to 2 d.p. standard deviation = 2.298 to 3 d.p.

10 mean = 9 standard deviation = 1.6073

11 mean = 0.85 standard deviation = 1.21 to 2 d.p.
mode = 0 range = 5

12 mean = 1.75 standard deviation = 1.09 to 2 d.p.

13 mean = 13.83 to 2 d.p. standard deviation = 4.91313

14 All numbers in the data set are 95.

15 Set Z has the largest standard deviation and set X has the smallest.

16 No change to standard deviation or variance.

17 mean = 8.964 st dev = 0.10346 var = 0.01