Dispersion and Variability: Exercises and Solutions

Measures of dispersion and variability of data

Dispersion refers to the variability or spread in the data. The most important measures of dispersion are

(1) the ***average deviation***,

(2) **the *variance***, and

(3) the ***standard deviation***.

We will measure these for populations and samples, as well as for grouped and ungrouped data.

1 . The **Average Deviation** (*AD*)

(a) for **ungrouped** data is given by

$$AD = \frac{\sum |X - \mu|}{N}$$ for populations and

$$AD = \frac{\sum |X - \bar{X}|}{n}$$ for samples

Where the two vertical bars indicate the absolute value, where the arithmetic mean, *or average*, of a population is represented by μ (the Greek letter μ); and for a sample, by $\bar{X}$ (read X bar). For **ungrouped data**, μ and $\bar{X}$ are calculated by the following formulas:

$$\mu = \frac{\sum X}{N} \quad \text{and} \quad \bar{X} = \frac{\sum X}{n}$$

where $\sum X$ refers to the sum of all the observations, while N and n refer to _the number of observations in the population and sample, respectively. For **grouped** data μ and $\bar{X}$ are calculated by

$$\mu = \frac{\sum fX}{N} \quad \text{and} \quad \bar{X} = \frac{\sum fX}{n}$$

Example 1 :A storeowner took an inventory count and found that he had sold 124 dresses during Spring Break holidays. 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 and the AD?

Solution:

This is an example of a discrete series where the classes in the distribution consist of single values. Do the computation as follows:

Size, x	Number of dresses, f	$f\,x$
4	3	12
5	10	50
6	18	108
7	20	140
8	24	192
9	20	180
10	14	140
11	10	110
12	5	60
Total	124	992

Multiple the x number by the f number on the same line and place the result in the third column. Example, **7 × 20 = 140**

$\sum f$

$\sum f \cdot x$

$$\mu = \frac{\sum (f \cdot x)}{\sum f} = \frac{992}{124} = 8$$ The mean size of the dress is 8.

4	3	8	-4	4	12
5	10	8	-3	3	30
6	18	8	-2	2	36
7	20	8	-1	1	20
8	24	8	0	0	0
9	20	8	1	1	20
10	14	8	2	2	28
11	10	8	3	3	30
12	5	8	4	4	20

Average Deviation = 196/124 =1.580645

Total	124				196

where $\sum fX$ refers to the sum of frequency of each class *f* times the class midpoints *X*. Therefore, the **average deviation** (AD) for **grouped** data is given by

$$AD = \frac{\sum f\left|X - \mu\right|}{N}$$ for populations and

$$AD = \frac{\sum f\left|X - \overline{X}\right|}{n}$$ for samples

2. Midpoint Measurement

Each interval is characterized by its **midpoint**, 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:

$$\text{midpoint} = \frac{\text{Low} + \text{High}}{2}$$

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

So the final table looks like this:

Midpoints (x)	Frequency (f)
2.5	0
7.5	9
12.5	13
17.5	1

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

Example 2

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:
Set up the computations as in the following table.

Class Interval	Midpoint X	f	fx	$f.x^2$
20 and under 25	22.5	3	67.5	1518.75
25 and under 30	27.5	5	137.5	3781.25
30 and under 35	32.5	9	292.5	9506.25
35 and under 40	37.5	4	150.0	5625.00
40 and under 45	42.5	3	127.5	5418.75
Total		24	775.0	25 850.00
		↑ N	↑ $\sum f x$	↑ $\sum f x^2$

$$\sigma^2 = \frac{\sum f.x^2}{N} - \left(\frac{\sum f.x}{N}\right)^2 = \frac{25\,850}{24} - \left(\frac{775}{24}\right)^2 = 34.331\,597\,23$$

= 1,077.083 – (600,625/576) = 1,077.083 – 1,042.75 = 34.331

	A	B	C	D
1	**Lower Limit**	**Upper Limit**	**Mid value**	**Frequency**
2	**20**	**25**	**22.5**	**3**
3	**25**	**30**	**27.5**	**5**
4	**30**	**35**	**32.5**	**9**
5	**35**	**40**	**37.5**	**4**
6	**40**	**45**	**42.5**	**3**

1. Type the words “Variance” in cell A8.

2. Highlight cell B8 (B8 will hold the answer for the variance).

3. Select the „insert function“ key ***fx.***

3. Standard Deviation,

The population standard deviation σ and sample standard deviation s are the positive square roots of their respective variances. For **ungrouped** data is given by

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}}$$ for populations and

$$S = \sqrt{\frac{\sum (X-\overline{X})^2}{n-1}}$$ for samples

For **grouped** data is given by

$$\sigma = \sqrt{\frac{\sum f(X-\mu)^2}{N}}$$ for populations and

$$S = \sqrt{\frac{\sum f(X-\overline{X})^2}{n-1}}$$ for samples

The most widely used measure of (absolute) dispersion is the standard deviation. Other measures (besides the variance and average deviation) are the *range*, **the inter-quartile range**, and **the quartile deviation**.

***4.* The coefficient of variation** *V*

measures **relative** dispersion:

$$V = \frac{\sigma}{\mu}$$ for populations and

$$V = \frac{s}{\overline{X}}$$ for samples

Example 3

Adam received the following grades (measured from 0 to 10) on the 10 quizzes he took during a semester: 6, 7, 6, 8, 5, 7, 6, 9, 10, and 6. These grades can be arranged into frequency distributions as in Table 1 and shown graphically as in Fig. 1.

Table 1 Frequency Distributions of Grades

Grades	Absolute Frequency	Relative Frequency
5	1	0.1
6	4	0.4
7	2	0.2
8	1	0.1
9	1	0.1
10	1	0.1
	10	1.0

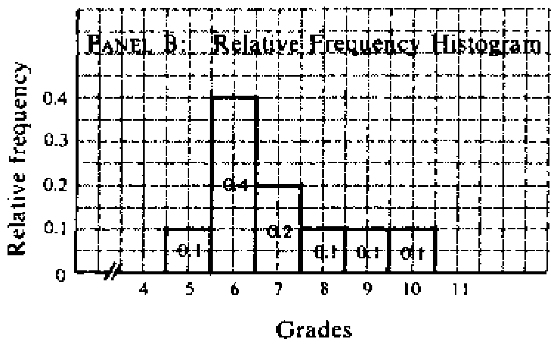

Figure 1

Example 4

The average deviation, variance, standard deviation, and coefficient of variation for the ungrouped data given in Example 1 ($\mu = 7$) can be found with the aid of the following table

Table 2 Calculations on the Data in Example 1

Grade *X*	μ	$X - \mu$	$\lvert X - \mu \rvert$	$(X - \mu)^2$
6	7	− 1	1	1
7	7	0	0	0
6	7	− 1	1	1
8	7	1	1	1

5 7 6 9 10 6	7 7 7 7 7 7	-2 0 -1 2 3 -1	2 0 1 2 3 1	4 0 1 4 9 1
		$\sum(x-\mu)=0$	$\sum\lvert x-\mu\rvert=12$	$\sum(x-\mu)^2=22$

Example 5

The cans in a sample of 20 cans of fruit contain net weights of fruit ranging from 19.3 to 20.9 oz, as following:

19.7, 19.9, 20.2, 19.9, 19.9, 20.0, 20.6, 19.3, 20.4, 19.9, 20.3, 20.1, 19.5, 20.9, 20.3, 20.8, 19.9, 20.0, 20.6, 19.9, 19.8.

If we want to group these data into 6 classes, we get **class intervals** of 0.3 oz [(21.0 - 19.2)/6 = 0.3 oz]. The weights given above can be arranged into the frequency distributions given in Table 3:

The average deviation, variance, standard deviation, and coefficient of variation for the frequency distribution of weights (grouped data) given in Table 2.3 can be found with the aid of Table 3 (V = 20.08 oz).

where $$AD = \frac{\sum f\lvert X-\mu\rvert}{N} = \frac{6.36}{20} \approx 0.318$$

$$S^2 = \frac{\sum f(X-\bar{X})^2}{n-1} = 2.9520/19 = 0.1554 \text{ oz squared},$$

$$S = \sqrt{2.952/19} \approx 0.394,\ V = S/\bar{X} \approx 0.0196 \text{ or } 1.96\ \%.$$

Note that in the formula for S^2 and S, $n - 1$ rather than n is used in the denominator. From the formulas for σ^2, σ, s^2, and s given in this section, others may be derived that will simplify the calculations for a large body of data.

Table 3 Frequency Distribution of Weights

Weight, oz	Class Midpoint	Absolute Frequency	Relative Frequency	Cumulative Frequency
19.2-19.4	19.3	1	0.05	1
19.5-19.7	19.6	2	0.10	3
19.8-20.0	19.9	8	0.40	11
20.1-20.3	20.2	4	0.20	15
20.4-20.6	20.5	3	0.15	18
20.7-20.9	20.8	2	0.10	20
		= 20	= 1. 00	

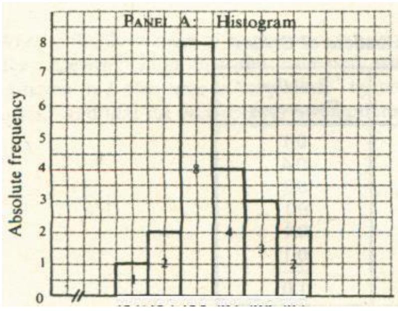

Figure 2

<u>Range</u>

The range of a quantitative data is the difference between the largest and smallest values of a series.
Range = Maximum – Minimum

Example 6

Find the standard deviation of the weekly salaries of the Jones family.

Weekly salaries in dollars are: 151 140 313 164 90

Solution:

1. Your first step is to calculate the arithmetic mean.

$$\mu = \frac{151 + 140 + 313 + 164 + 90}{5} = 171.6$$

2. Build a table that services the standard deviation formula.

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

x	$x-\mu$	$(x-\mu)^2$
151	151 – 171.6 = –20.6	424.36
140	140 – 171.6 = –31.6	998.56
313	313 – 171.6 = 141.4	19 993.96
164	164 – 171.6 = –7.6	57.76
90	90 – 171.6 = –81.6	6658.56

$\sum (x-\mu)^2$ = 28 133.2

Plug into the above formula and evaluate:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}} = \sqrt{\frac{28133.2}{5}} = 75.01093254 = 75$$

Having obtained the standard deviation, what does the number tell you?

We know that the standard deviation is a measure of variability, and that the bigger the value of the standard deviation is, the more variability we have in the set of data.

Example 7

Find the variance of the data given in example 23.

Solution:

The standard deviation in example 2 was determined to be $\sigma = 75.01093254$

Variance is defined as $\sigma^2 = (\sigma)^2 = (75.01093254)^2 = 5626.64001$

Example 25

Find the variance of the ABC Bank customer waiting times. The times in minutes are:

6.5 6.6 6.7 6.8 7.1 7.3 7.4 7.7 7.7 7.7

Solution:

1. Your first step is to calculate the arithmetic mean.

$$\mu = \frac{6.5 + 6.6 + 6.7 + 6.8 + 7.1 + 7.3 + 7.4 + 7.7 + 7.7 + 7.7}{10} = 7.15 \text{ minutes}$$

2. Build a table that services the variance formula.

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

x	$x - \mu$	$(x - \mu)^2$
6.5	–0.65	0.4225
6.6	–0.55	0.3025
6.7	–0.45	0.2025
6.8	–0.35	0.1225
7.1	–0.05	0.0025
7.3	0.15	0.0225
7.4	0.25	0.0625
7.7	0.55	0.3025
7.7	0.55	0.3025
7.7	0.55	0.3025
		$\sum (x - \mu)^2 = 2.045$

Substitute into the above formula and evaluate:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} = \frac{2.045}{10} = 0.2045$$

Excel Solution

Open an Excel sheet.

Key in all ten numbers in cells A1:A10.
Highlight all the cells A1:A20 and sort it from A-Z↓ botton

Highlight cell A12 Write in the cell A12 =VAR(A1:A11)
Click OK. Write the number you get, should be = 2.045

To find out which bank is more consistent in its services evaluate the standard deviation **or** the variance for each bank. The lower the measure of variability is, the more consistent the bank is.

	Standard deviation	Variance
ABC Bank	0.452216762	0.2045
Zoho Bank	1.728149299	2.9865

Both the standard deviation and the variance of ABC Bank are smaller than those of Zoho Bank. This leads to the conclusion that ABC Bank has less variability than Zoho Bank, and hence it is more consistent.