Section 6.3 Measuring Spread

Question 1 – What is the range of the dataset?

Question 2 – What is the variance and standard deviation of a dataset?

Question 1 – What is the range of the dataset?

Key Terms

Range

Summary

The range is the simplest way to measure how spread out a set of data is. It is calculated by taking the highest data value and subtracting the lowest data value,

$$\text{Range} = \text{Maximum data value} - \text{Minimum data value}$$

Notes

Guided Example

Suppose the grades on an exam for a class are recorded in the table below.

55	65	75	85	95

Find the range of the exam scores.

Solution The lowest score is 55 and the highest score is 95. This makes the range

$$\text{Range} = 95 - 55 = 40$$

Practice

1. Suppose the grades on an exam for a class are recorded in the table below.

73	74	75	76	77

Find the range of the exam scores.

Question 2 – What is the variance and standard deviation of a dataset?

Key Terms

Variance Standard deviation

Summary

Variance and standard deviation are measures of how spread out a set of data is. If the variance is calculated from a population, the population variance σ^2 (sigma squared) of data x_i is the mean of the squared deviations,

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

where μ is the population mean and N is the population size.

The population standard deviation σ is the square root of the population variance,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

The sample variance s^2 and sample standard deviation s are calculated with similar formulas,

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} \qquad s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

where $\overline{x}$ is the sample mean and n is the sample size. The formulas are almost identical with μ corresponding to $\overline{x}$ and N corresponding to n. The major difference is in the denominator of the fraction. For population statistics, divide by N. For sample statistics, divide by $n - 1$.

To calculate the standard deviation by hand, it is useful to use a table like the one below.

x	$x-\overline{x}$	$(x-\overline{x})^2$
35		
32		
41		
28		
31		

The data values 35, 32, 41, 28, and 31 have been entered in the first column. To calculate how far each of these data lie from the mean, calculate the sample mean

$$\bar{x} = \frac{35+32+41+28+31}{5} = 33.4$$

In the second column of the table, subtract the mean from each of the data values.

x	$x-\bar{x}$	$\left(x-\bar{x}\right)^2$
35	1.6	
32	-1.4	
41	7.6	
28	-5.4	
31	-2.4	

To fill out the third column, square each of the entries in the second column and place the result in the third column.

x	$x-\bar{x}$	$\left(x-\bar{x}\right)^2$
35	1.6	2.56
32	-1.4	1.96
41	7.6	57.76
28	-5.4	29.16
31	-2.4	5.76

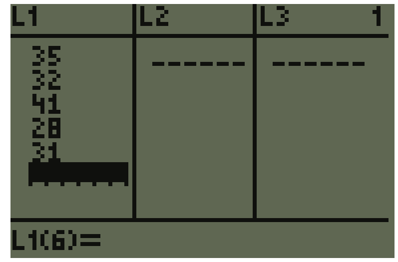
To calculate the sum $\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$, sum the entries in the third column:

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 = 2.56+1.96+57.76+29.16+5.76 = 97.2$$

Finishing the calculation, we get

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}} = \sqrt{\frac{97.2}{5-1}} \approx 4.93$$

On a graphing calculator, we can calculate the standard deviation by following the steps below:

1. Start by entering the data into a list on the calculator. Press [STAT][ENTER] to enter a list. 2. Enter each data value under L1 followed by [ENTER]. 3. Once each data value has been entered, press [2nd][MODE] to quit the list editor and return to the home screen.	
4. From the home screen, press [2nd][STAT] to access the LIST commands. 5. Press [▶][▶] to move the cursor to the MATH menu. 6. Press [▼][▼][▼][▼][▼][▼][ENTER] or [7] to paste the stdDev(command to the home screen.	
7. After the stdDev(command, we need to insert the name of the list we are using. Press [2nd][1][)] to paste the name of the list, L1 and a parentheses into the homescreen. 8. Press [ENTER] to calculate the median.	

<u>Notes</u>

Guided Example

Suppose the grades on an exam for a class are recorded in the table below.

55	65	75	85	95

a. Find the variance of the exam scores. Assume these data values are a sample of a larger set of data.

Solution Start by calculating the mean,

$$\bar{x} = \frac{55+65+75+85+95}{5} = 75$$

Now fill out a table to help you calculate the value of $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.

x	$x - \bar{x}$	$(x - \bar{x})^2$
55	-20	400
65	-10	100
75	0	0
85	10	100
95	20	400

Add the numbers in the third column and divide by one less than the sample size to get the variance

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1000}{4} = 250$$

b. Find the standard deviation of the exam scores. Assume these data values are a sample of a larger set of data.

Solution The standard deviation is the square root of the variance,

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{250} \approx 15.81$$

Practice

1. Suppose the grades on an exam for a class are recorded in the table below.

73	74	75	76	77

a. Find the variance of the exam scores. Assume these data values are a sample of a larger set of data.

b. Find the standard deviation of the exam scores. Assume these data values are a sample of a larger set of data.

Guided Example

Suppose an instructor records the score on an exam in a frequency table.

Class	Frequency
[40, 50)	1
[50, 60)	0
[60, 70)	5
[70, 80)	5
[80, 90)	9
[90, 100)	27

Using the midpoint of each class, estimate the standard deviation of the scores. Assume the scores are a sample.

Solution To estimate the standard deviation, we need to first calculate the mean using the frequencies,

$$\bar{x} = \frac{1\cdot 45 + 0\cdot 55 + 5\cdot 65 + 5\cdot 75 + 9\cdot 85 + 27\cdot 95}{47} \approx 86.70$$

When we create the table to calculate the standard deviation, we need to modify it slightly to utilize the frequencies:

x	$x-\bar{x}$	$\left(x-\bar{x}\right)^2$	$\left(x-\bar{x}\right)^2\cdot f$
45	-41.70	1738.89	1738.89
55	-31.70	1004.89	0
65	-21.70	470.89	2354.45
75	-11.70	136.89	684.45
85	-1.7	2.89	26.01
95	8.3	68.89	1860.03

The entries in the fourth column are found by multiplying the entries in the third column by the corresponding frequencies. Sum the entries in the fourth column to give

$$\sum\left(x-\bar{x}\right)^2\cdot f = 6663.83$$

Divide by one less than the sample size and take the square root to get the standard deviation,

$$s = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{n-1}} = \sqrt{\frac{6663.83}{46}} \approx 12.04$$

The frequencies are utilized so that instead of carrying out the sum over every data value, we simply sum over the different values (and multiply by the corresponding frequencies.

Practice

2. A sample of 20 college students are examined to determine the number of credit hours each student is taking. The results are summarized in the table below

Class	Frequency
[0, 6)	1
[6, 12)	3
[12, 18)	15
[18, 24)	1

Using the midpoint of each class, estimate the standard deviation of the scores. Assume the scores are a sample.