THE ERINDALE ACADEMY
1576 Dundas Street West, Mississauga ON L5C 1E5
www.teacademy.ca Tel: 905-232-1576
Email: info@teacademy.ca

Teacher: Ella

Unit 5: One and Two Variable data analysis

Lesson 5.5: Measures of spread – range, quartiles, percentiles

Learning Goal: analyze and describe data using statistical measures of spread

Range: a calculation of the _____ between the highest and lowers value in a data set.

- Gives no details as to how the data are spread out

Percentiles: divide the data into 100 intervals that have equal numbers of values. It is the percent of all the data in a set that are _____ a specific data value.

- Data must be sorted from lowest to highest
- k percent of the data are less than or equal to kth percentile, P_k
- and (100 – k) percent are greater than or equal to P_k
- e.g. Grade of Level 3 is 20th percentile which mean 20 percent of students are getting lower or equal level than level 3, and 80% of students are getting higher or equal level than 3.
- Standardized tests often use percentiles to convert raw scores to scores on a scale from 1 to 100.

Percentile rank calculates the value in a data set that falls in a given percentile.

, where p is the percentile, n is the size of the population, and R is the whole number rank of the data point. If R is not a whole number, round R down.

Percentile calculates the percent data that is less than or equal to a given value.

, where p is the percentile, n is the size of the population, L is the number of data less than the data value, and E is the number of data equal to the data value.

Example 1: The list below shows that marks from least to greatest for 25 students on a recent test out of 40.

a) Calculate the 80th percentile.
b) What percentile is a mark of 25?
c) What percentile is a mark of 40?

15	16	20	21	21
23	24	25	25	25
26	28	28	28	28
30	30	31	32	34
36	36	37	38	40

Teacher: Ella

Practice:

An audio magazine tested 60 different models of speakers and gave each one an overall rating based on sound quality, reliability, efficiency, and appearance. The raw scores for the speakers are listed in ascending order below.

35	47	57	62	64	67	72	76	83	90
38	50	58	62	65	68	72	78	84	91
41	51	58	62	65	68	73	79	86	92
44	53	59	63	66	69	74	81	86	94
45	53	60	63	67	69	75	82	87	96
45	56	62	64	67	70	75	82	88	98

a) If the Audio Maximizer Ultra 3000 scored at the 50th percentile, what was its raw score?

b) What is the 90th percentile for these data?

c) Does the SchmederVox's score of 75 place it at the 75th percentile?

d) What percentile is a mark of 71?

Quartiles: divides a set of ordered data into four groups with equal numbers of values.

- The first quartile, Q_1 – the median of the lower half of the data (also the 25gh percentile)
- The second quartile, Q_2 – the median of the entire data set (also the 50^{th} percentile)
- The third quartile, Q_3 – the median of the upper half of the data (also the 75^{th} percentile)
- If the number of data is even, take the midpoint between the two middle values as the median, Q_2
- If the number of data below the median is even, Q_1 is the midpoint between the two middle values in this half of the data
- Q_3 is determined in a similar way

Interquartile range: is the range for the middle half of the data or $Q_3 - Q_1$

- The larger the interquartile range, the larger the spread of the central half of the data

Semi-interquartile range: is one half of the interquartile range of $\frac{Q3 - Q1}{2}$.

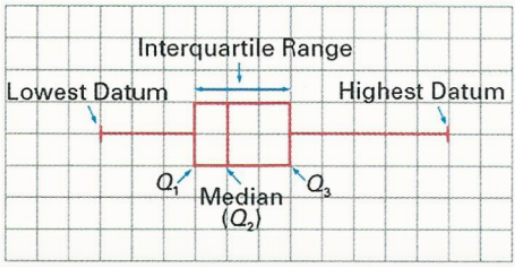
Both interquartile and semi-interquartile range both indicate how closely the data are clustered around the median.

Teacher: Ella

Example 2: a random survey of people at a science-fiction convention asked them how many times they have seen Star Wars. The results are shown below. Determine the median, the first and third quartiles, and the interquartile and semi-quartile ranges. What information do these measures provide?

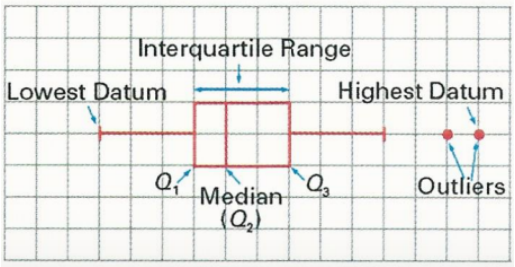
3, 4, 2, 8, 10, 5, 1, 15, 5, 16, 6, 3, 4, 9, 12, 3, 30, 2, 10, 7

Box-and-whisker-plot: is used to graphically illustrate the measurements represented in the quartiles.



Shows:

Modified Box-and-whisker-plot: is used when the data contain outliers. Usually, any point that is 1.5 times the box length away from the box is classified as an outlier. This method usually gives a clearer illustration of the distribution.



Shows:

Example 3: Using the Star War Survey data and calculations from example 2, create a suitable box plot.

Practise

A

1. Determine the mean, standard deviation, and variance for the following samples.
   a) Scores on a data management quiz (out of 10 with a bonus question):

5	7	9	6	5	10	8	2
11	8	7	7	6	9	5	8

   b) Costs for books purchased including taxes (in dollars):

12.55	15.31	21.98	45.35	19.81
33.89	29.53	30.19	38.20	

2. Determine the median, Q_1, Q_3, the interquartile range, and semi-interquartile range for the following sets of data.
   a) Number of home runs hit by players on the Statsville little league team:

6	4	3	8	9	11	6	5	15

   b) Final grades in a geography class:

88	56	72	67	59	48	81	62
90	75	75	43	71	64	78	84

3. For a recent standardized test, the median was 88, Q_1 was 67, and Q_3 was 105. Describe the following scores in terms of quartiles.
   a) 8
   b) 81
   c) 103

4. What percentile corresponds to
   a) the first quartile?
   b) the median?
   c) the third quartile?

5. Convert these raw scores to z-scores.

18	15	26	20	21

Apply, Solve, Communicate

B

6. The board members of a provincial organization receive a car allowance for travel to meetings. Here are the distances the board logged last year (in kilometres).

44	18	125	80	63	42	35	68	52
75	260	96	110	72	51			

   a) Determine the mean, standard deviation, and variance for these data.
   b) Determine the median, interquartile range, and semi-interquartile range.
   c) Illustrate these data using a box-and-whisker plot.
   d) Identify any outliers.

7. The nurses' union collects data on the hours worked by operating-room nurses at the Statsville General Hospital.

Hours Per Week	Number of Employees
12	1
32	5
35	7
38	8
42	5

   a) Determine the mean, variance, and standard deviation for the nurses' hours.
   b) Determine the median, interquartile range, and semi-interquartile range.
   c) Illustrate these data using a box-and-whisker plot.

8. Application
   a) Predict the changes in the standard deviation and the box-and-whisker plot if the outlier were removed from the data in question 7.
   b) Remove the outlier and compare the new results to your original results.
   c) Account for any differences between your prediction and your results in part b).

9. **Application** Here are the current salaries for François' team.

Salary ($)	Number of Players
300 000	2
500 000	3
750 000	8
900 000	6
1 000 000	2
1 500 000	1
3 000 000	1
4 000 000	1

a) Determine the standard deviation, variance, interquartile range, and semi-interquartile range for these data.

b) Illustrate the data with a modified box-and-whisker plot.

c) Determine the z-score of François' current salary of $300 000.

d) What will the new z-score be if François' agent does get him a million-dollar contract?

10. **Communication** Carol's golf drives have a mean of 185 m with a standard deviation of 25 m, while her friend Chi-Yan shoots a mean distance of 170 m with a standard deviation of 10 m. Explain which of the two friends is likely to have a better score in a round of golf. What assumptions do you have to make for your answer?

11. Under what conditions will Q_1 equal one of the data points in a distribution?

12. **a)** Construct a set of data in which $Q_1 = Q_3$ and describe a situation in which this equality might occur.

b) Will such data sets always have a median equal to Q_1 and Q_3? Explain your reasoning.

13. Is it possible for a set of data to have a standard deviation much smaller than its semi-interquartile range? Give an example or explain why one is not possible.

14. **Inquiry/Problem Solving** A business-travellers' association rates hotels on a variety of factors including price, cleanliness, services, and amenities to produce an overall score out of 100 for each hotel. Here are the ratings for 50 hotels in a major city.

39	50	56	60	65	68	73	77	81	87
41	50	56	60	65	68	74	78	81	89
42	51	57	60	66	70	74	78	84	91
44	53	58	62	67	71	75	79	85	94
48	55	59	63	68	73	76	80	86	96

a) What score represents

i) the 50th percentile?

ii) the 95th percentile?

b) What percentile corresponds to a rating of 50?

c) The travellers' association lists hotels above the 90th percentile as "highly recommended" and hotels between the 75th and 90th percentiles as "recommended." What are the minimum scores for the two levels of recommended hotels?

ACHIEVEMENT CHECK

Knowledge/ Understanding	Thinking/Inquiry/ Problem Solving	Communication	Application

15. a) A data-management teacher has two classes whose midterm marks have identical means. However, the standard deviations for each class are significantly different. Describe what these measures tell you about the two classes.

b) If two sets of data have the same mean, can one of them have a larger standard deviation and a smaller interquartile range than the other? Give an example or explain why one is not possible.