Measures of Variability

Purpose:
Participants will determine the variance, standard deviation, range, and interquartile range of a data set.

Overview:
In pairs, participants will input data into the lists of their graphing calculator. They will determine the variance, standard deviation, range, and interquartile range of the data and provide possible interpretations for these measures of variability.

TExES Mathematics 4-8 Competencies. The beginning teacher:

IV.012.B	Applies concepts of center, spread, shape, and skewness to describe a data distribution.
IV.012.D	Demonstrates an understanding of measures of central tendency (e.g., mean, median, mode) and dispersion (e.g., range, interquartile range, variance, standard deviation).
IV.012.E	Analyzes connections among concepts of center and spread, data clusters and gaps, data outliers, and measures of central tendency and dispersion.
IV.012.F	Calculates and interprets percentiles and quartiles.

TEKS Mathematics Objectives. The student is expected to:

5.13.B	Describe characteristics of data presented in tables and graphs including the shape and spread of the data and the middle number.
6.10.B	Use median, mode, and range to describe data.
7.12.A	Describe a set of data using mean, median, mode, and range.
7.12.B	Choose among mean, median, mode, or range to describe a set of data and justify the choice for a particular situation.

Terms.
mean, first quartile, third quartile, spread of data, variance, standard deviation, range, interquartile range

Materials.
For instructor:

- Transparencies
- Overhead graphing calculator and LCD panel

For each participant:

- Data Sheet
- Activity Sheet
- Graphing calculator

Transparencies.

- *Measures of Variability*

Activity Sheet(s).

- *Measures of Variability*

Procedure:

Steps	Questions/Math Notes
1. Have participants enter the data either by hand or by linking two calculators and sending the lists from the calculator which has the data to the other calculator. Ask participants to work in pairs and to check one another's input of data.	Let L1 = Region Number, L2 = Total Number of Students Enrolled in that Region during 1987-88, and L3 = Total Number of Students Enrolled in that Region during 1997-98. Do not enter the 4^{th} column of data.
2. Once the data for L1, L2, and L3 are entered into all calculators, have participants compute the total student enrollment change from 1987-88 to 1997-98 in L4. Once L4 has been calculated, have participants exchange calculators with their partner and check the L4 column for correctness. If L4 is correct, then most likely all the data is correct.	Calculating the total student enrollment change from 1987-88 to 1997-98 in L4 can be done by placing the cursor on L4 and typing L3 – L2. Then press ENTER.
3. Ask participants to calculate the sample variance and sample standard deviation of L3.	*What information do we need in order to calculate the sample variance of L3?* (We need to know the data values and the mean of L3.) *Once we know the sample mean, then what do we do?* (We find $(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, $(x_3 - \bar{x})^2$, $(x_4 - \bar{x})^2$, …, $(x_n - \bar{x})^2$). *Then what do we do?* (We find the sum of the terms $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2$ and then divide the sum by n-1 or 19 in this case.) *What does this number give us?* (It gives us the sample variance of the data.) *How do we determine the sample standard deviation?* (We take the square root of the sample variance to determine the sample standard deviation.) *What does the sample standard deviation tell us in this problem?* (See Solution section below for possible responses.)

4. Ask participants to calculate the range of the data in L3.	*How do we calculate the range of the data?* (We need the data listed in ascending order so we can determine the minimum and maximum data values. Then we subtract min X from max X.) *What does the range tell us in this problem?* (See Solution section below for possible responses.)
5. Ask participants to calculate the interquartile range of the data in L3.	*How do we calculate the interquartile range of the data in L3?* (First we need to determine the values of the 1st quartile and 3rd quartile.) *Then what do we do?* (We calculate Q3 – Q1.) *What does the interquartile range tell us in this problem?* (See Solution section below for possible responses.)

Calculator Instructions.

<u>Entering Data in Lists</u>. We begin by entering the data into the calculator. On a TI-83 calculator, press STAT, 1: Edit, and enter data.
L1 = Texas Education Region Numbers
L2 = Total Students Enrolled in Region 1987-88
L3 = Total Students Enrolled in Region 1997-98

<u>Transferring Data from One Calculator to Another</u>. Use the TI calculator cable to link two calculators together. For the calculator which contains the data, press 2nd LINK, SEND, 4: List, SELECT, and then highlight L1, L2, and L3.
For the calculator which is going to receive the data, press 2nd LINK, RECEIVE, ENTER.
Once the receiving calculator is ready to receive the data, press TRANSMIT on the sending calculator.

<u>Computing the Sample Variance</u>. The initial step in computing the sample variance of L3 is to calculate:
$(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, $(x_3 - \bar{x})^2$, $(x_4 - \bar{x})^2$, …, $(x_n - \bar{x})^2$. We can do this by placing the cursor on L5 and typing (L3 – 2nd LIST, MATH, 3: mean(L3)) ^{2}.
Next we need to sum the data in L5, so, 2nd Quit to get back to the home screen, 2nd LIST, MATH, 5: sum(L5), ENTER.
Then we need to divide the sum by (n – 1), which is 19.

Sum(L4) = 7.57649868E11
Ans/19 = 3.987630884E10 or approximately 39,876,308,840

The variance of the data listed in L3 is 3.987630884E10.

Computing the Sample Standard Deviation. To compute the sample standard deviation of L3, we simply take the square root of the sample variance.

Another way to obtain the sample standard deviation of L3 you can simply press 2nd LIST, MATH, 7: stdDev(L3).

The sample standard deviation of the data listed in L3 is approximately 199,690 students.

Computing the Range. To compute the range of the data listed in L3, we simply subtract the smallest data value from the largest. When the data is listed in ascending or descending order, these values are easy to locate. If the data is not listed in ascending or descending order, then we can ask the calculator for the minimum and maximum values of the list. Press 2nd LIST, MATH, max(L3), ENTER and the calculator will inform you that the maximum value of the list is 828,262. Press 2nd LIST, MATH, min(L3), ENTER, and the calculator will inform you that the minimum value of the list is 42,388.

Range = max X – min X = 828,262 (Region IV) – 42,388 (Region IX) = 785,874 students

Computing Quartiles. The 25th percentile is called the first quartile of x and the 75th percentile is the third quartile of x. The first quartile value of L3 is an average of the 5th and 6th data entries when the data are listed in ascending order. Hence, the first quartile value is the mean of 57,730 (Region III) and 80,711 (Region XVI) or 69,220.5 total students. The third quartile value of L3 is an average of the 15th and 16th data entries when the data are listed in ascending order. Hence, the third quartile value is the mean of 247,989 (Region XIII) and 284,614 (Region I) or 266,301.5 total students.

Q1 = 69,220.5
Q3 = 266,301.5
Of course, 0.5 of a student isn't a reasonable answer.

We can also determine the first and third quartile values using the calculator. Press STAT, CALC, 1: 1-Var Stats L3, ENTER. This command provides the mean value ($\bar{x}$), the sum of the x values ($\sum x$), the sum of the x^2 values ($\sum x^2$), the sample standard deviation, the population standard deviation, value of n, minimum value of x, Q1, Median, Q3, and maximum value of x.

Computing the Interquartile Range. To compute the interquartile range, subtract Q1 from Q3. Interquartile Range = Q3 – Q1 = 266,301.5 – 69,220.5 = 197,081 students.

Solutions to Activity Sheet:

1. The sample variance of L3 = 3.987630884E10 or approximately 39,876,308,840 students.

2. The sample standard deviation of L3 = approximately 199,690 students. The sample standard deviation tells us that this data is greatly dispersed. This means that the number of students enrolled in the education service regions in Texas during 1997-98 varied greatly.

3. The range of L3 = max X – min X = 828,262 (Region IV) – 42,388 (Region IX) = 785,874 students. The range tells us that there is a large dispersion between the minimum and maximum values. Region IV (Houston) served many more students than Region IX (Wichita Falls) during the 1997-98 school year. Even though Region IX appears to cover more square miles of land than Region IV, the population of Region IX is much less dense.

4. The interquartile range = Q3 – Q1 = 266,301.5 – 69,220.5 = 197,081 students. The interquartile range tells us that even the data values near the center point are widely dispersed. There is a nearly 200,000 student population difference between the middle 50% of the data.