Characterizing Variation in Data

IQs, Variance, and Std.Dev.
MDM 4U0
Data is Usually Spread Out

- And if you're going to use a MoCT to describe where the data cluster, you should also use a number to tell people how spread out the data is.
- Why not show the distribution?

A MoCT would be near the peak of the mounded distribution, but the data is fairly spread out [dcn.ed.ac.uk].
Spread/Dispersion/Variation

- Can be characterized by:
  - Interquartile range (IQR)
  - Variance
  - Standard Deviation

[vosesoftware.com].
Interquartile ranges

- Sort the data (least to greatest) and split the data into quarters.
- The median $= Q_2$.
- The median of the half of the data with smaller values $= Q_1$.
- The median of the other half of the data $= Q_3$.
- The IQR $= Q_3 - Q_1$. 

Q also stands for quality [aeservices.ca]
Interpreting Qn and IQR

- Q1 contains the lower ____% of data.
- Q2 contains the lower ____% of data.
- Q3 contains the lower ____% of data.
- The IQR contains ____% of the data.

Q1, Q2, Q3, and the IQR illustrated for a data set [answers.com].
Which One (Q1, Q2, Q3, IQR)

- Measures the **spread** of data?

Q1, Q2, Q3, and the IQR illustrated for a data set [answers.com].
An Example

- Final marks for two years of a class. Calculate all Qn and the IQR. Which class was “better”?
  - Year 1: 68, 40, 74, 40, 76, 78, 93, 89, 20, 75, 55, 96, 83, 67, 85, 57, 88, 90
  - Year 2: 63, 73, 73, 87, 80, 93, 87, 74, 77, 80, 85, 52, 62, 80, 84, 72, 72, 67

Q also stands for quality [aeservices.ca]
Variance

- Go back to our mark set from the other day: 10/20, 3/5, 33/55, 8/10.
- What was the mean? Deviation? Average deviation?
- What's the IQR?

[neos.mcs.anl.gov]
Deviation Consternation

- The average deviation for a data set is zero (if your MoCT is a simple mean).
- But deviation is a natural measure of spread.
- How could we make deviations always positive?
Variance

- **Variance**: the average of the sum of the squares of the deviations.

- Write this in summation notation.

A conjecture. Can you make it a proof? [physicsforums.com]
Variance

- Your book uses $\sigma^2$ for variance. Generally, $\sigma$ should be used for a population; $s$ should be used for a sample of a population.

\[ \text{Var}(x) = s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \]

Variance written in summation notation.
Standard Deviation

- Is the **square root of the variance**.
- Write this in summation notation.
- Why the square root?
- This is really the **sample standard deviation**, but your text calls it the standard deviation.

[decodeunicode.com]
(Sample) Standard Deviation

- Again, your book also uses $\sigma$, but should really use $s$.

The sample standard deviation written in summation notation.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$
And Again With the Marks

- 10/20, 3/5, 33/55, 8/10
- Calculate the variance and standard deviation. Compare them with the IQR.

Two distributions with the same mean, but different SDs. If you're going to state A MoCT, you better state one of the MsoVs [JRBrown, PD].
Assumption(s)

- What assumption(s) are there in the SD?
- We may or may not get to fixing this in the course.
Your Calculator

- If it has an SD or STAT mode, can help you.
- You enter data, then use the mean button and/or the SD button. *Et voila* – only data entry is required.

The only permitted calculator in Chem 232 at Uvic [web.uvic.ca]
With Calculators

- Go back to the two years' of class marks. Work out the mean, variance, and sample standard deviation using SD or STAT mode on your calculator.

The only permitted calculator in Chem 232 at Uvic [web.uvic.ca]
The Class Marks, Again

- Year 1: 68, 40, 74, 40, 76, 78, 93, 89, 20, 75, 55, 96, 83, 67, 85, 57, 88, 90
- Year 2: 63, 73, 73, 87, 80, 93, 87, 74, 77, 80, 85, 52, 62, 80, 84, 72, 72, 67

Marks being celebrated [travelpod.com].
Grouped Data

- You can simplify manual calculations (and some calculator calculations) if your data can be grouped.
- What's the MsoCT and MsoV for the following data? 1, 1, 2, -1, 1, 0, 1

\[ s = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}{n}} \]
You Can't Run Away From Proof

- Show that
- Hint: start with the regular expression for $s^2$. 

\[ s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - n\bar{x}^2 \]
Proof

A proof is a convincing demonstration [washingtonpost.com].
Two distributions with the same mean, but different SDs. If you're going to state A MoCT, you better state one of the MsoVs [JRBrown, PD].
Which?

- There are two datasets below. One is from a real class; the other is a set of numbers from a simple random sampling (1+99*ran), pegged according to board rules. Which is which? Justify your answer with a numerical index.
  - 40, 62, 72, 62, 20, 20, 58, 40, 20, 98, 65, 72, 40, 65, 72, 68
  - 58, 40, 62, 20, 58, 68, 62, 62, 78, 62, 52, 72, 68, 65, 65

[servitokss.com]