MTH 23 LECTURE NOTES (Ojakian)

Topic 4: Variance, Standard Deviation

OUTLINE

References (**Algebra Book**: 8.2, 8.4; **Statistics Book**: 2.7)

1. Intuition and calculation of Variance and Standard Deviation
2. Chebyshev's Theorem

---

1. Motivation and Ideas
   (a) How "spread out" is the data? Some Examples ...
      i. 4, 8, 5, 4, 7, 8
      ii. 9, 9, 9, 9, 9
      iii. 1, 5, 87, 2, 4, 40
      iv. Our class' height versus our class' age. Which do yo expect to be more spread out? Why?
   (b) Typical measures of spread: Variance and Standard Deviation
      i. Real numbers versus Integers.
      ii. Each is a real number that is zero or larger.
      iii. The larger it is, the more spread out the data is.
   (c) A detail: Sample versus Population
      i. s^2 = Sample Variance
      ii. s = Sample Standard Deviation
      iii. σ^2 = Population Variance
      iv. σ = Population Standard Deviation
2. Using Excel
   (a) In Excel (apply to above data):
      i. **Sample** Variance: `VAR.S`
      ii. **Sample** Standard Deviation: `STDEV.S`
      iii. **Population** Variance: `VAR.P`
      iv. **Population** Standard Deviation: `STDEV.P`
   (b) A class game
      i. In turn each person in class will pick a real number
      ii. You will see the current mean and current standard deviation
      iii. Your goal is that in the end (when the last student announces their number), the mean is as close to 100 as possible, with a standard deviation as close to 5 as possible.

3. How to calculate the Variance and Standard Deviation

PROBLEM 1. *Consider the data: 2, 3, 3, 8, 10, 10 (from Example 6 in 5th Edition)*

By hand, using the worksheet, showing all the steps, *calculate the Population Variance, the Population Standard Deviation, the Sample Variance, and the Sample Standard Deviation.*

(a) Find the **mean**

(b) Find the **differences** from the mean

(c) **Square** the differences

(d) **Sum** these square differences

(e) **Divide** appropriately:

i. By the number of data (if it is population data)

ii. By the number of data, MINUS ONE (if it is sample data)

(f) Now you have the **variance**!

(g) To get the **standard deviation,** take the **square root**.

4. Some background mathematics: Radicals

(a) Recall square roots

PROBLEM 2. *Calculate the following:* $\sqrt{16}, \sqrt{100}$

PROBLEM 3. *Which of the following "make sense":* $\sqrt{16}, \sqrt{-16}, \sqrt{0}$

(b) Inverses

PROBLEM 4.

i. Pick your favorite number: square it, then take its square root. What happened?

ii. Now try reversing the order in the last two parts, taking the root first, then the power. What happened?

iii. Express these mathematical points using mathematical expressions.

PROBLEM 5. *Where do you see something like these inverses in the calculation of standard deviation?*

PROBLEM 6. *Note: Use Excel Worksheet for this problem.*

Consider with examples why in the calculation of variance/SD, replacing squaring by cubing, or doing nothing, is a bad idea? (For cubing, pick values symmetric around 0)

5. Variance and Standard Deviation ... Again!

PROBLEM 7. *Suppose you have a data set which consists of the same number V repeated. What is its mean and standard deviation?*

PROBLEM 8. *From section 3.2 (8th edition) do: 7*

PROBLEM 9. *If the sample variance is* 50 *calculate the exact sample standard deviation. If the population standard deviation is* $\sqrt{23}$ *calculate the exact population variance.*

PROBLEM 10. *Suppose I gave some data and we calculate the mean to be 70 and the standard deviation to be 25. Suppose we are using sample data.*

*(***Note: Do similar modifications to data from last question.***)*

(a) *What happens to the mean if some 73s are put in the data?*

(b) *What happens to the mean if some 70s are put in the data?*

(c) *What happens to the standard deviation if some 70s are put in the data?*

(d) *What happens to the standard deviation if some 73s are put in the data?*

PROBLEM 11. *Consider some student data from our class.* **Using the Excel worksheet, showing all the steps**, *calculate the Population Variance, the Population Standard Deviation, the Sample Variance, and the Sample Standard Deviation.*

6. Chebyshev's Theorem

Theorem 1. *Consider some data that has some mean and some standard deviation. You choose any real number $k > 1$, and let $C = 1 - \frac{1}{k^2}$ (write as a percent). Then the following is true:*

> *Within k standard deviations of the mean, we have at least C percent of the data.*

PROBLEM 12. *Play the Chebyshev game using Excel.*

PROBLEM 13. *Consider some class data. Use Excel to find the sample mean and sample standard deviation. Determine a Chebyshev interval about the mean in which at least 75% of the data fall. Verify it.*

PROBLEM 14. *Do the last problem again, but now determine a Chebyshev interval about the mean in which at least 88.9% of the data fall. Verify it.*

How would you do it for a percent different from $75\%, 88.9\%, 93.8\%$? ...