Section 3.2 Measure of Variation

Objective: In this lesson you learned to find range, variance, and standard deviation; to compute the coefficient of variation; and to apply Chebyshev's theorem to raw data.

Important Vocabulary

Range | Variance | Standard Deviation | Sum of Squares | Sample Variance

Sample Standard Deviation | Population mean | Population Variance

Population Standard Deviation | Coefficient of Variation | Chebyshev's Theorem

Important Notation

s^2 s μ σ^2 σ CV

The **range** is: the difference between the largest and smallest data values

Focus Point:
- Find the range, variance, and standard deviation

I. Variance and Standard Deviation

What are two measures that will measure the distribution or spread of data around an expected value (either $\bar{x}$ or μ)?

Variance and Standard Deviation

Quantity	Formula	Description
x		the variable x represents a data value or outcome.
Mean	$\bar{x} = \frac{\Sigma x}{n}$	the average of the data, or what you "expect" to happen next time
	$x - \bar{x}$	the difference between what did happen and what you expected to happen; this is a "deviation"
	$\sum (x-\bar{x})^2$	the sum of squares.

Sum of Squares	Formula	Description
	$\sum (x-\bar{x})^2$	the defining formula
	$\sum x^2 - \frac{(\sum x)^2}{n}$	the computation formula

* Both give same answer *

Sample Variance	Formula	Description
defining formula	$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$	The sample variance is s^2
computation formula	$s^2 = \frac{\sum x^2 - (\sum x)^2/n}{n-1}$	• think of this as an average of the $(x-\bar{x})^2$ values.

Sample Standard Deviation	Formula	Description
defining formula	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$	This is sample standard deviation, s.
computation formula	$s = \frac{\sum x^2 - (\sum x)^2/n}{n-1}$	• a measure of risk • the larger the s-value, the larger the risk.

FORMULAS - SAMPLE STATISTICS

	Sample Variance	Sample Standard Deviation
Defining	$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$
Computation	$s^2 = \frac{\sum x^2 - (\sum x)^2/n}{n-1}$	$s = \sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n-1}}$

x = data value
$\bar{x}$ = mean

n = # of data values

What do measures of variation tell us?

- Range – difference between largest and smallest data values.
- Standard deviation – an average of data spread around the mean
- Variance – square of standard deviation; also a measure of spread around the mean.

Population Parameters

mean: $\mu = \frac{\Sigma x}{N}$

Variance: $\sigma^2 = \frac{\Sigma (x-\mu)^2}{N}$

Standard deviation: $\sigma = \sqrt{\frac{\Sigma (x-\mu)^2}{N}}$

II. Coefficient of Variation

What is a disadvantage of the standard deviation as a unit of measure?

depends on the units of measure.

How is the *coefficient of variation* expressed?

as a percentage of the sample/population mean

Focus Point:

- Compute the coefficient of variation from raw data. Why is the coefficient of variation important?

If $\bar{x}$ and s represent: the sample mean and standard deviation, then the Coefficient of Variation (CV) is

$$C.V. = \frac{s}{\bar{x}} * 100\% \quad \text{or} \quad C.V. = \frac{\sigma}{\mu} * 100\%$$

Population form

σ = pop. STD. DEV.

μ = pop. mean

III. Chebyshev's Theorem

Chebyshev's Theorem

For any set of data, and any constant $k > 1$, the proportion of data that must lie within k standard deviations on either side of the mean is at least $1 - \frac{1}{k^2}$

Focus Point:

- Apply Chebyshev's theorem to raw data. What does a Chebyshev interval tell us?

Results of Chebyshev's Theorem

For any set of data:

- at least 75% from $\mu - 2\sigma$ to $\mu + 2\sigma$
- at least 88.9% from $\mu - 3\sigma$ to $\mu + 3\sigma$
- at least 93.8% from $\mu - 4\sigma$ to $\mu + 4\sigma$

* $\bar{x}$ and s can be used in place of μ and σ *

What does Chebyshev's Theorem tell us?

Applies to ANY distribution

- minimum percentage of data falls between the mean and any specified number of standard deviations above/below.

- A minimum of 88.9% of the data falls between 3 standard deviations above/below the mean.

Section 3.2 Examples – Measures of Variation

(1) Big Blossom Greenhouse gathered another random sample of mature peak blooms from Hybrid B. The six blossoms had the following widths (in inches):

5 5 5 6 7 8

a. Again, we will construct a table so that we can find the mean, variance, and standard deviation more easily. In this case, what is the value of n? Find the sum of Column I in the table, and compute the mean.

$\bar{x} = \frac{36}{6} = 6$

I	II
x	x^2
5	25
5	25
5	25
6	36
7	49
8	64
$\sum x =$ 36	$\sum x^2 =$ 224

b. Complete Column II in the table.

c. What is the value of n? of $n - 1$? Use the computation formula to find the sample variance s^2. *NOTE:* Be sure to distinguish between $\sum x^2$ and $(\sum x)^2$.

$n = 6$
$n - 1 = 5$

$$s^2 = \frac{\sum x^2 - (\sum x)^2/n}{n-1} = \frac{224 - (36)^2/6}{5} = 1.6$$

d. Use a calculator to find the square root of the variance. Is this the standard deviation?

$s = \sqrt{1.6} \approx 1.26$

this is the standard deviation

(2) Cabela's in Sindey, Nebraska, is a very large outfitter that carries a broad selection of fishing tackle. It markets its products nationwide through a catalog service. A random sample of 10 spinners (a type of fishing reel) from Cabela's extensive spring catalog gave the following prices (in dollars):

1.69 1.49 3.09 1.79 1.39 2.89 1.49 1.39 1.49 1.99

a. Use a calculator with sample mean and sample standard deviation keys to compute $\bar{x}$ and s.

$\bar{x} = \$1.87$

$s = \$0.62$

b. Compute the CV for the spinner prices at Cabela's.

$$C.V. = \frac{s}{\bar{x}} \times 100\% = \frac{\$0.62}{\$1.87} \times 100\% = \boxed{33.2\%}$$

(3) The *East Coast Independent News* periodically runs ads in its classified section offering a month's free subscription to those who respond. In this way, management can get a sense about the number of subscribers who read the classified section each day. Over a period of 2 years, careful records have been kept. The mean number of responses per ad is $\bar{x} = 525$ with standard deviation $s = 30$.

Determine a Chebyshev interval about the mean in which at least 88.9% of the data fall.

$\bar{x} - 3s$ to $\bar{x} + 3s$

$525 - 3(30)$ to $525 + 3(30)$

$\boxed{435 \text{ to } 615}$