PROGRAMME: M.Ed (I Semester)
PROGRAMME CODE: 2PAEDU
COURSE: Introduction to Educational Research

COURSE CODE: CC-3
UNIT : V
TOPIC: Measures of Variability

Measures of Variability

Dr.I. Muthuchamy
Professor
Department of Educational Technology
Bharathidasan University
Tiruchirappalli-620 023

Measures of Variability

Measures of variability tell us:

- The extent to which the **scores differ from each other** or **how spread out** the scores are.

- How **accurately** the **measure of central tendency describes** the distribution.

- The **shape** of the distribution.

Measures of Variability

Some Definitions:

Population (universe) is the collection of things under consideration

Sample is a portion of the population selected for analysis

Statistic is a summary measure computed to describe a characteristic of the sample

Measures of Variability

Just what is variability?

Variability is the **spread or dispersion of scores**.

Measuring Variability

There are a few ways to measure variability and they include:

1) **The Range**
2) **The Mean Deviation**
3) **The Standard Deviation**
4) **The Variance**

Variability

Measures of Variability

Range: The range is a measure of the distance between highest and lowest.

R= H – L

Temperature Example:

		Range:
Chennai:	**89° – 65°**	**24°**
Hyderabad:	**106° – 41°**	**65°**

The Range

The simplest and most straightforward measurement of variation is the range which measures variation in interval-ratio variables.

Range = highest score – lowest score

$$R = H - L$$

Range: Examples

If the **oldest** person included in a study was 89 and the **youngest** was 18, then the *range* would be 71 years.

Or, if the **most** frequent incidences of disturbing the peace among 6 communities under study is 18 and the **least** frequent incidences was 4, then the *range* is 14.

Okay, so now you tell me the range…

This table indicates the number of areas, as defined by the Census Bureau, in seven Places.

What is the range in the number areas in these seven place?

- R=H-L
- R=9-3
- R=6

Chennai	**3**
Kolkata	**4**
New Delhi	**4**
Mumbai	**5**
Bangalore	**4**
Hyderabad	**3**
Trivandrum	**9**

Variance and Standard Deviation

Variance: is a **measure of the dispersion of a sample** (or how closely the observations cluster around the mean [average]). Also known as the **mean of the squared deviations.**

Standard Deviation: the **square root of the variance**, is the measure of variation in the observed values (or variation in the clustering around the mean).

The Variance

Remember that the **deviation** is the distance of any given score from its mean.

$$(X - \overline{X})$$

The **variance** takes into account **<u>every</u>** score.

But if we were to simply add them up, the plus and minus (positive and negative) scores would cancel each other out because the sum of actual deviations is always zero!

$$\sum (X - \overline{X}) = 0$$

The Variance

So, what we should we do?

We **square** the actual deviations and then **add** them together.

$$\sum (X - \bar{X})^2$$

- Remember: When you square a negative number it becomes positive!

SO,

S^2 = sum of squared deviations divided by the number of scores.

The variance provides information about the relative variability.

Variance: Weeks on Unemployment:

Step 1: Calculate the Mean | Step 2: Calculate Deviation | Step 3: Calculate Sum of square Dev

X (weeks)	Deviation: $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)		
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$		
ΣX=30 $X = \frac{30}{6} = 5$		$\sum (X - \bar{X})^2 = 52$		

The Variance

The mean of the squared deviations is the same as the variance, and can be symbolized by s^2

$$s^2 = \frac{\sum\left(X - \overline{X}\right)^2}{N}$$

where $s^2 =$ variance

$\sum(X - \overline{X})^2 =$ sum of the squared deviations from the mean

$N =$ total number of scores

Variance: Weeks on Unemployment:

Step 1: Calculate the Mean

Step 2: Calculate Deviation

Step 3: Calculate Sum of square Dev

Step 4: Calculate the Mean of squared dev.

X (weeks)	**Deviation:** $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)	**Variance:** $s^2 = \frac{\sum (X - \bar{X})^2}{N}$	
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$	$\frac{52}{6} = 8.67$	
ΣX=30 $\bar{X} = \frac{30}{6} = 5$		$\sum (X - \bar{X})^2 = 52$		

What is a standard deviation?

Standard Deviation:

It is the typical **(standard)** difference **(deviation)** of an observation from the mean.

Think of it as the average distance a data point is from the mean, although this is not strictly true.

What is a standard deviation?

Standard Deviation:

The standard deviation is calculated by taking **the square root of the variance**.

$$s = \sqrt{\sum \frac{(X - \bar{X})^2}{n}}$$

Variance: Weeks on Unemployment:

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev	Step 4: Calculate the Mean of squared dev.	Step 5: Calculate the Square root of the Var.

X (weeks)	**Deviation:** $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)	**Variance:** $s^2 = \frac{\sum (X - \bar{X})^2}{N}$	**Standard Deviation:** $s = \sqrt{\sum \frac{(X - \bar{X})^2}{n}}$ (square root of the variance)
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$	$\frac{52}{6} = 8.67$	$\sqrt{8.67}$
ΣX=30 $\bar{X} = \frac{30}{6} = 5$		$\sum (X - \bar{X})^2 = 52$		s = 2.94

Variance: Weeks on Unemployment:

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev	Step 4: Calculate the Mean of squared dev.	Step 5: Calculate the Square root of the Var.
X (weeks)	**Deviation:** $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)	**Variance:** $s^2 = \frac{\sum (X - \bar{X})^2}{N}$	**Standard Deviation:** s =√Σ(X - χ)2 / N (square root of the variance)
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$	$\frac{52}{6} = 8.67$ (weeks squared)	$\sqrt{8.67}$
ΣX=30 χ= $\frac{30}{6}$=5		$\sum (X - \bar{X})^2 = 52$		s = 2.94

Two Ways to Look at Standard Deviation: Sample and Population

Standard Deviation

- **Standard deviation** is the positive square root of the variance.
  - Population Standard Deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

  - Sample Standard Deviation:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}}$$

G. Baber, Department of Statistics
University of South Carolina; Slide 44

The Variance

S^2 = sum of squared deviations divided by the number of scores.

Some Things to Remember about Sample Variance

Uses the deviation from the mean

Remember, the sum of the deviations always equals zero, so you have to square each of the deviations

S^2= sum of squared deviations divided by the number of scores

Provides information about the relative variability

What is a standard deviation?

It is the typical **(standard)** difference **(deviation)** of an observation from the mean.

Think of it as the average distance a data point is from the mean, although this is not strictly true.

Two Ways to Look at Standard Deviation: Sample and Population

Standard Deviation

- **Standard deviation** is the positive square root of the variance.
  - Population Standard Deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

  - Sample Standard Deviation:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}}$$

G. Baber, Department of Statistics
University of South Carolina; Slide 44

Variability

Weeks on Unemployment: Standard Deviation (S)

Thus, in this example, "the distribution deviates from the mean by nearly **3 weeks**." In other words, on average…the scores in this distribution **deviate from the mean by nearly 3 weeks**."

Judges and Sentencing:

You are an attorney with a choice of judges:

Judge A: gives all defendants convicted of assault **24 months**

Judge B: gives all defendants convicted of assault **6-months to 6 years** depending on priors.

Prison Sentences: Judge A

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev	Step 4: Calculate the Mean of squared dev.	Step 5: Calculate the Square root of the Var.
X (months)	$X - \chi$	**Deviation:** $(X - \chi)^2$ (raw score from the mean, squared)	**Variance:** $S^2 = \frac{\Sigma(X - \chi)^2}{N}$ ("the mean of the squared deviation.")	**Stand. Deviation:** $S = \sqrt{\frac{\Sigma(X - \chi)^2}{N}}$ (square root of the variance)
34 30 31 33 36 34	34-33= 1 30-33=-3 31-33=-2 33-33=0 36-33=3 34-33=1	$1^2 = 1$ $-3^2 = 9$ $-2^2 = 4$ $0^2 = 0$ $3^2 = 9$ $1^2 = 1$	$\frac{24}{6} = 4.0$ (months squared)	$\sqrt{4.0}$
$\Sigma X=198$ $\chi = \frac{198}{6} = 33$		$\Sigma(X - \chi)^2 = 24$		S = 2.0

Prison Sentences: Judge B

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev	Step 4: Calculate the Mean of squared dev.	Step 5: Calculate the Square root of the Var.
X (months)	X - χ	**Deviation:** $(X - \chi)^2$ (raw score from the mean, squared)	**Variance:** $S^2 = \frac{\Sigma(X - \chi)^2}{N}$ ("the mean of the squared deviation.")	**Stand. Deviation:** $S = \sqrt{\frac{\Sigma(X - \chi)^2}{N}}$ (square root of the variance)
26 43 22 35 20 34	36-30= -6 43-30=13 22-30=-8 35-30=5 20-30=-10 34-30=4	$-6^2 = 36$ $13^2 = 169$ $-8^2 = 64$ $5^2 = 25$ $10^2 = 100$ $4^2 = 8$	$\frac{402}{6} = 67.0$ (months squared)	$\sqrt{67.0}$
$\Sigma X = 180$ $\chi = \frac{180}{6} = 30$		$\Sigma(X - \chi)^2$ =402		S = 8.18

Variability

Judges and Sentences:

Judge A has a larger mean, smaller variance and standard deviation. "Judge A seems harsher but fairer. Judge B is more lenient but more inconsistent."

Thus Judge A would be a better judge to face. Their mean is larger (so you will get a longer sentence) but are less likely to get longer sentence than if you faced Judge B.

Standard Deviation

Is the square root of the variance.

Uses the same units of measurement as the raw scores.

Tells how much scores deviate below and above the mean

Standard Deviation

Standard Deviation: Applications

Standard deviation also allows us to:

1) Measure the **baseline of a frequency polygon.**
2) Find the **distance between raw scores and the mean** – a standardized method that permits comparisons between raw scores in the distribution – as well as between different distributions.

Standard Deviation

Standard Deviation: Baseline of a Frequency Polygon.

The baseline of a frequency polygon can be measured in units of standard deviation.

Example:

$\bar{X}$ = 80

s = 5

Then, the raw score 85 lies one Standard Deviation above the mean (+1s).

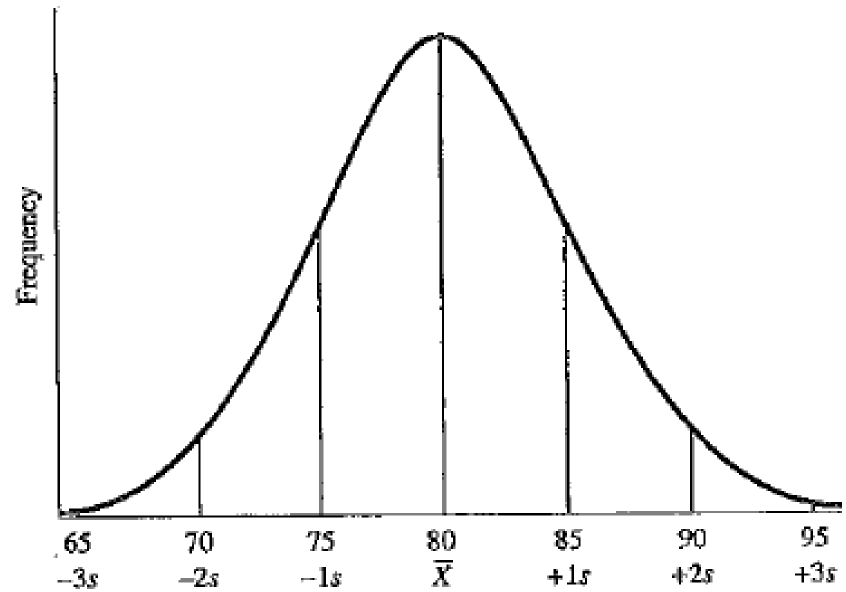


FIGURE 4.2 *Measuring the Base Line in Units of Standard Deviation When the Standard Deviation (s) Is 5 and the Mean ($\bar{X}$) Is 80*

Standard Deviation

Standard Deviation: The Normal Range

Unless highly skewed, approximately **two-thirds of scores** within a distribution will fall within the **one standard deviation above and below the mean**.

Example: Reading Levels

Words per minute.

$\bar{X}$ = 120

s = 25

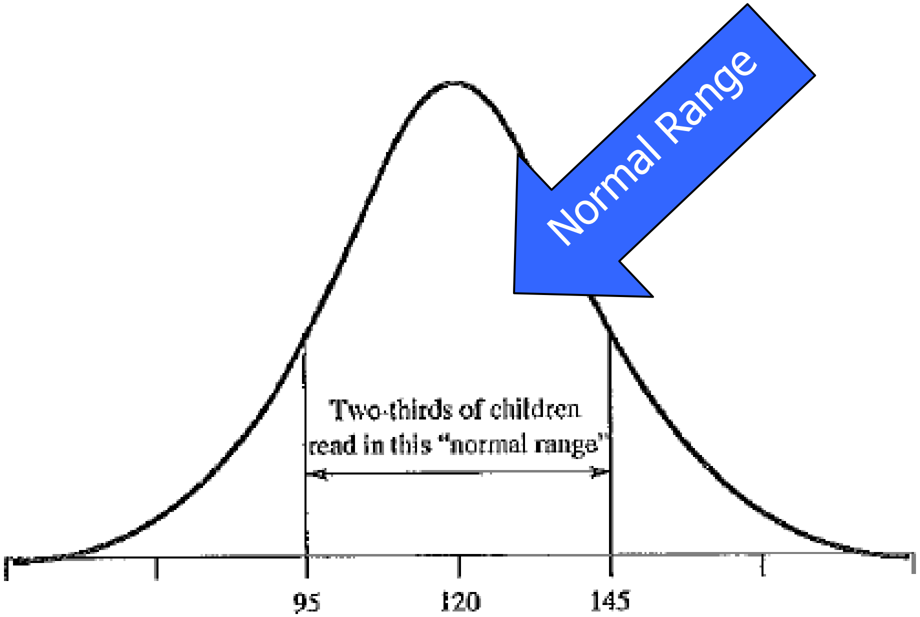


FIGURE 4.3 *Distribution of Reading Speed*

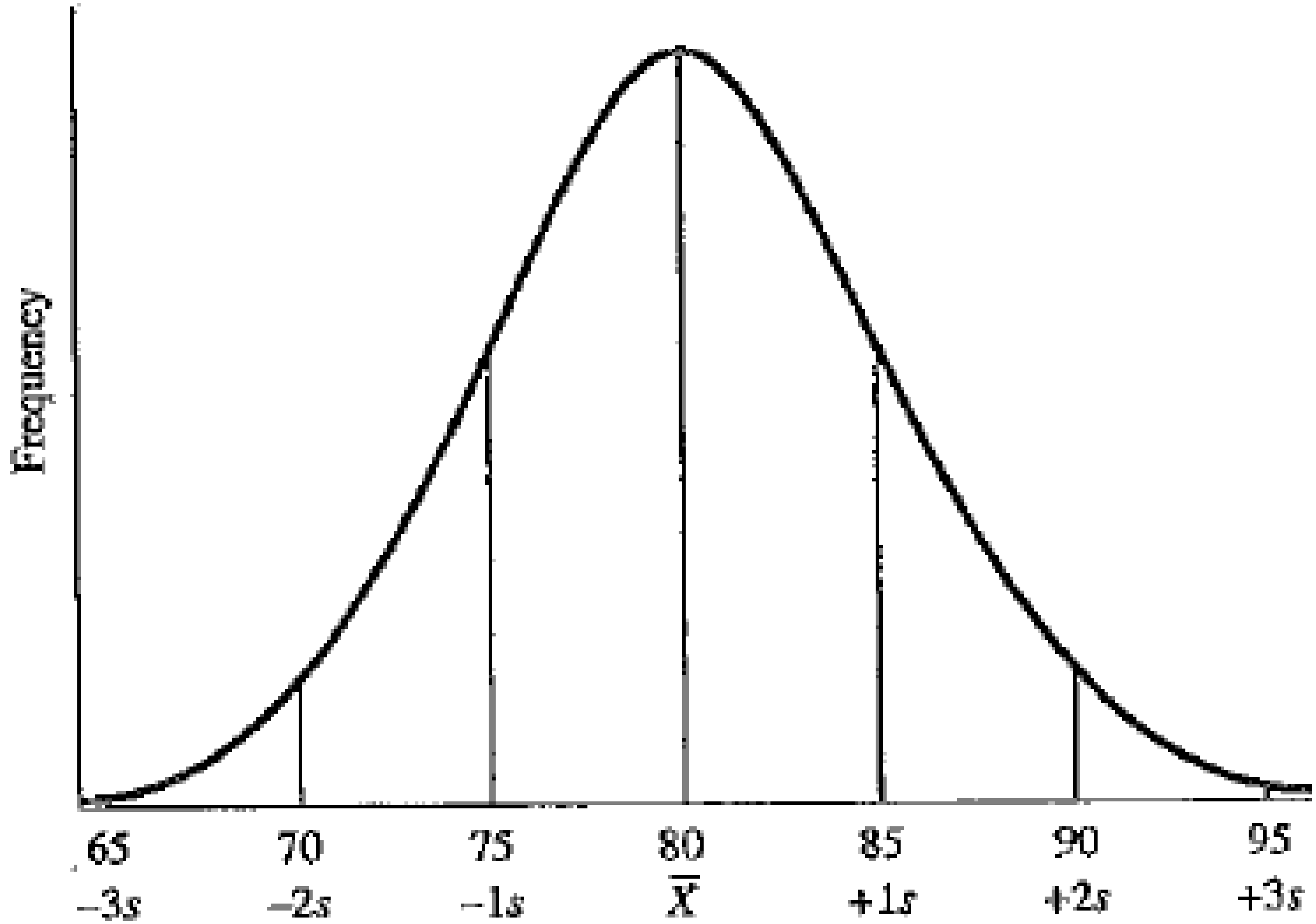
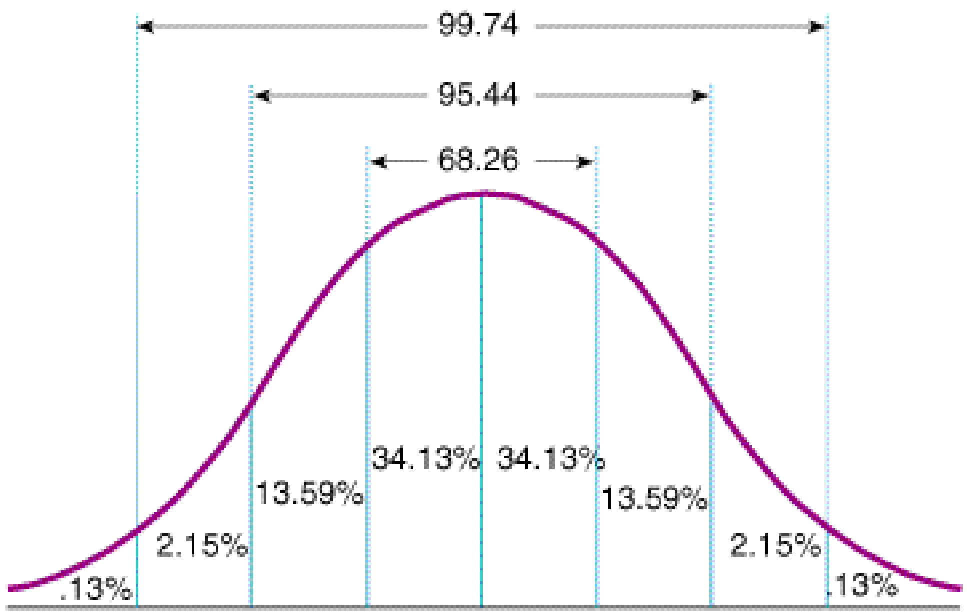
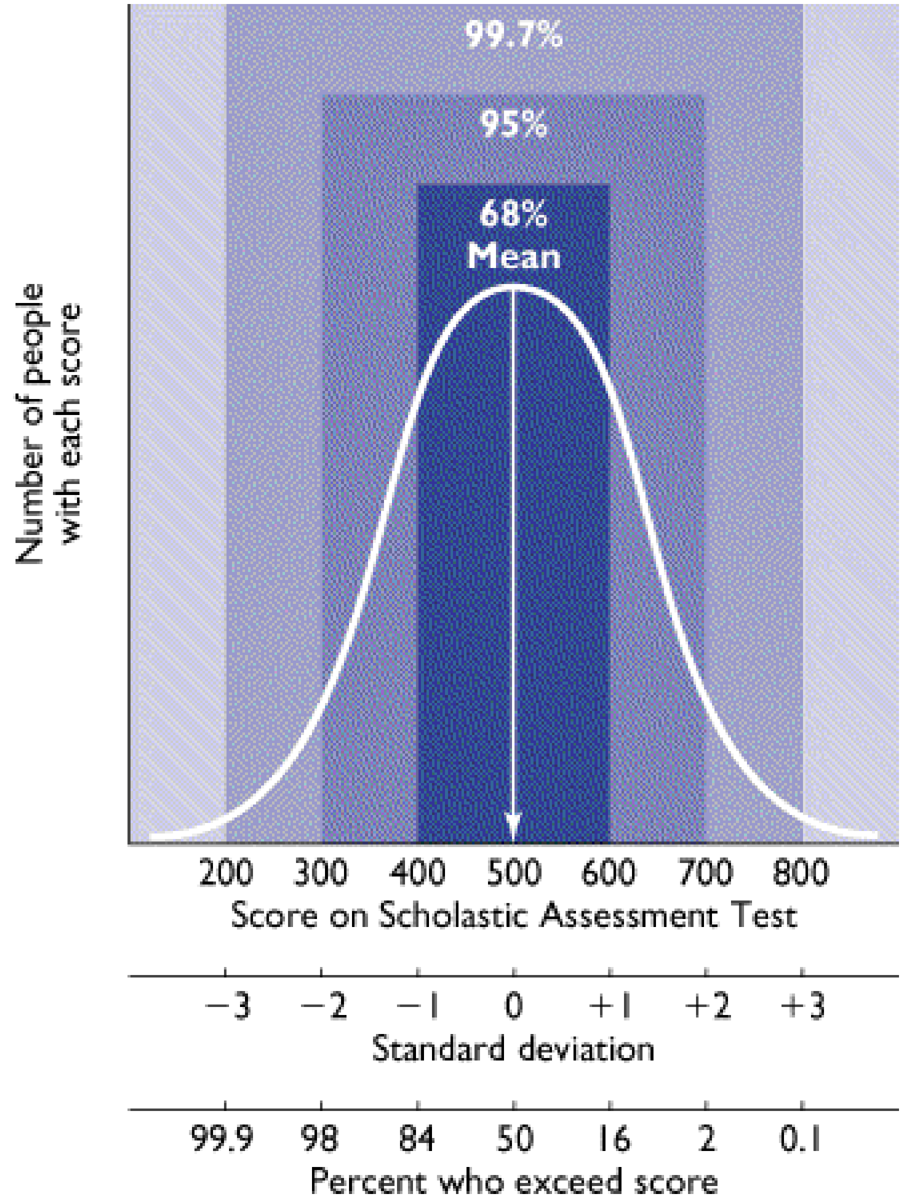


FIGURE 4.2 ***Measuring the Base Line in Units of Standard Deviation When the Standard Deviation (s) Is 5 and the Mean ($\overline{X}$) Is 80***



Standard Deviations	–3	–2	–1	0	+1	+2	+3
Cumulati Percentages	.1	2.3	15.9	50	84.1	97.7	99.9
Z Scores	%	%	%	%	%	%	%
T Scores	–3.	–2.	–1.0	0	+1.0	+2.0	+3.0
SAT Scores	0	0	40	50	60	70	80
Spelling Scores	20	30	400	500	600	700	800
IQ Scores	20	300	8	11	14	17	20


99.7%
95%
68%
Mean
Number of people with each score
200 300 400 500 600 700 800
Score on Scholastic Assessment Test
−3 −2 −1 0 +1 +2 +3
Standard deviation
99.9 98 84 50 16 2 0.1
Percent who exceed score

Standard Deviation

The deviation (definitional) formula for the population standard deviation

$$\sigma = \sqrt{\frac{\sum\left(X - \overline{X}\right)^2}{N}}$$

The larger the standard deviation the more variability there is in the scores

The standard deviation is somewhat less sensitive to extreme outliers than the range (as N increases)

Standard Deviation

The deviation (definitional) formula for the **sample** standard deviation

$$S = \sqrt{\frac{\sum\left(X - \overline{X}\right)^2}{N}}$$

What's the difference between this formula and the population standard deviation?

In the first case, all the *X*s represent the entire population. In the second case, the *X*s represent a sample.

Standard Deviation

Calculating S using the Raw-Score Formula

$$S = \sqrt{\frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N}}$$

To calculate ΣX^2 you square all the scores first and then sum them

To calculate $(\Sigma X)^2$ you sum all the scores first and then square them

The Raw-Score Formula: Example

X	X^2
21	441
25	625
24	576
30	900
34	1151
$X = 134$	$X^2 = 3698$

$$S = \sqrt{\frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N}}$$

$$S = \sqrt{\frac{3698 - \frac{134^2}{5}}{5}}$$

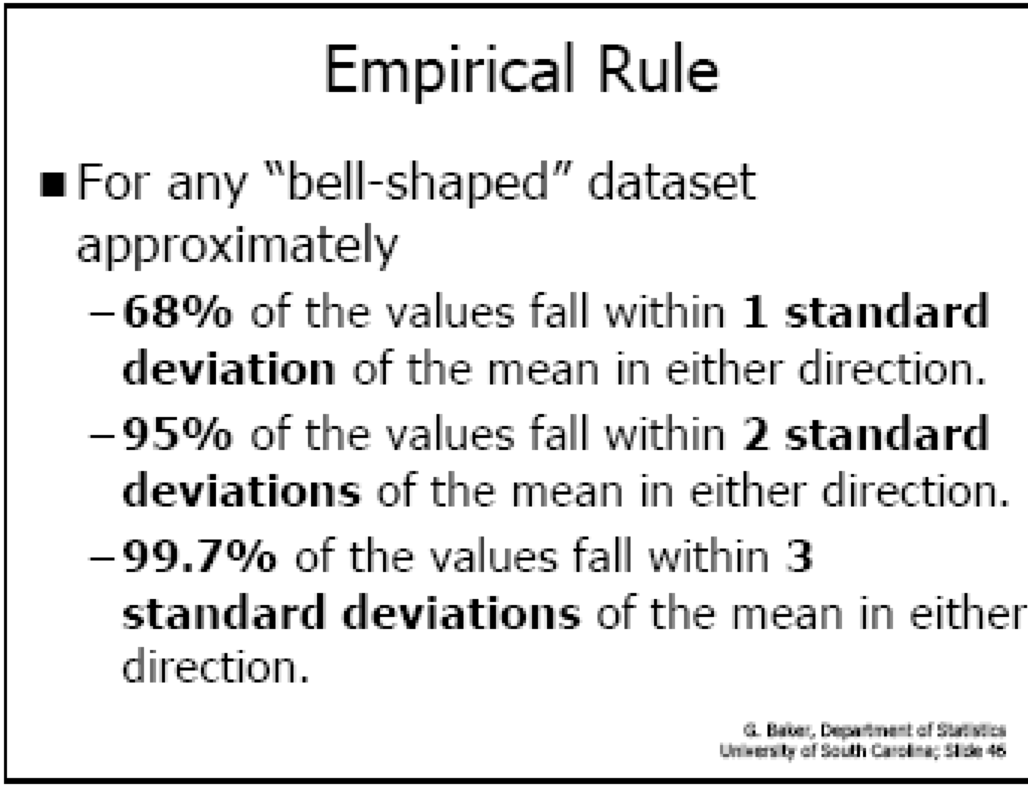
$$S = \sqrt{\frac{3698 - 3591.2}{5}}$$

$$S = \sqrt{21.36} = 4.62$$

Application to Normal Distribution

Knowing the standard deviation you can describe your sample more accurately

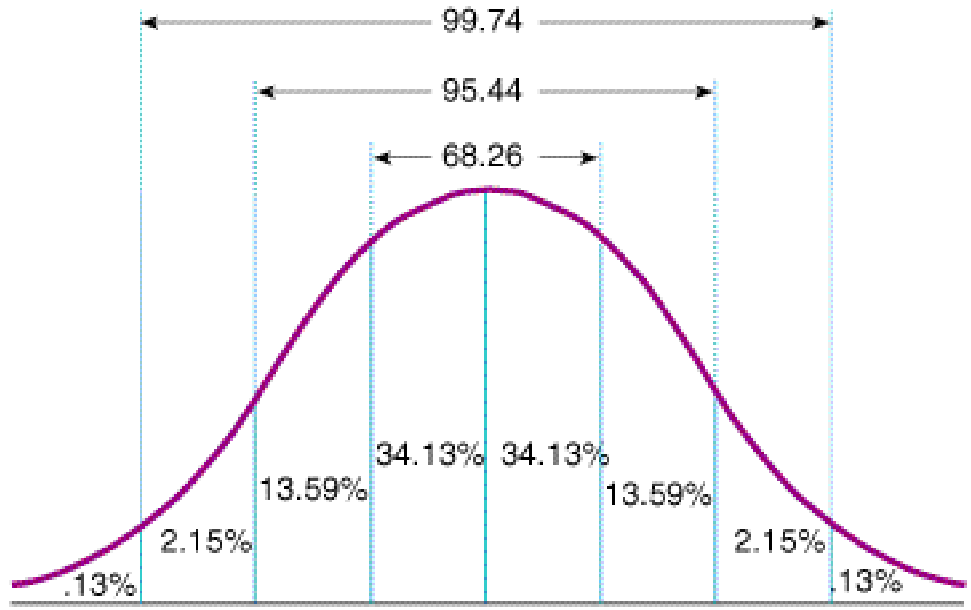
Look at the inflection points of the distribution



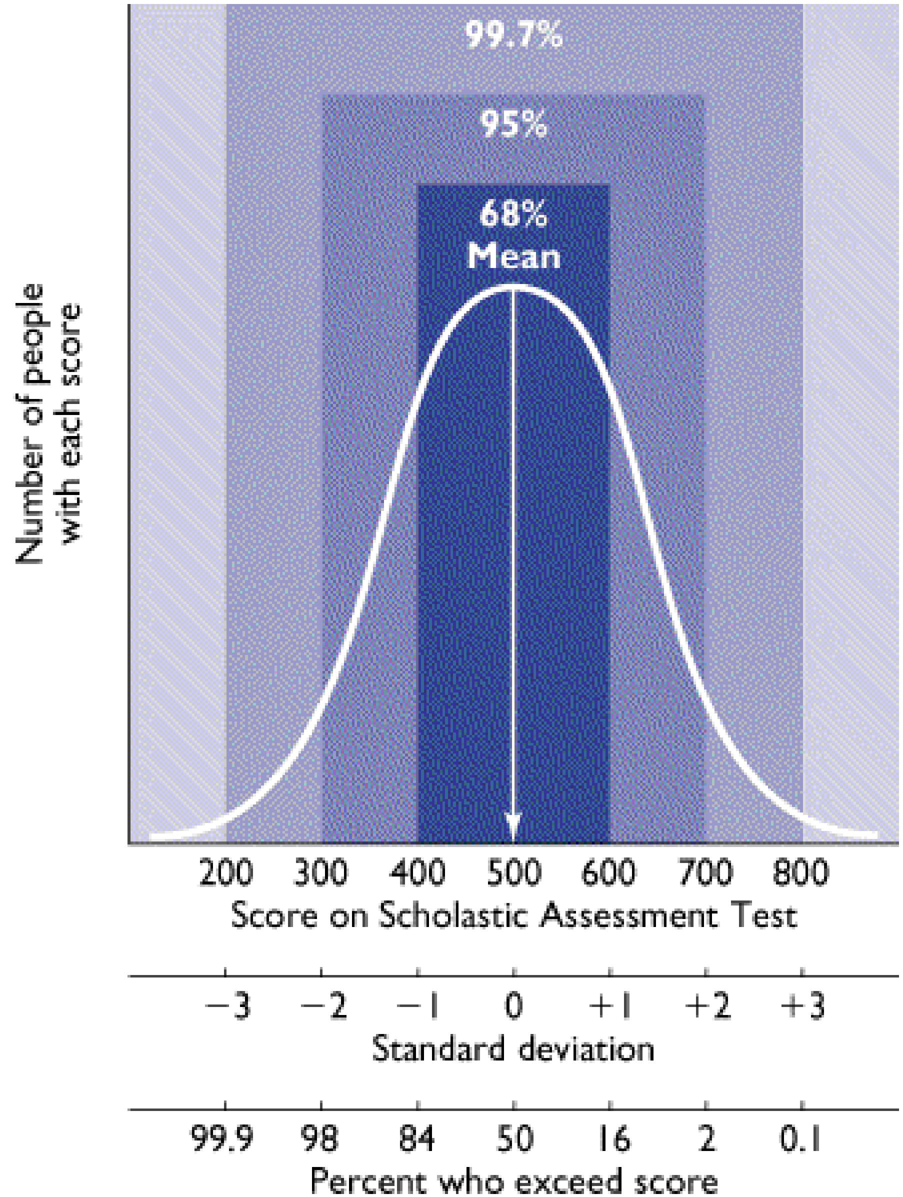
Empirical Rule

- For any "bell-shaped" dataset approximately
  - **68%** of the values fall within **1 standard deviation** of the mean in either direction.
  - **95%** of the values fall within **2 standard deviations** of the mean in either direction.
  - **99.7%** of the values fall within **3 standard deviations** of the mean in either direction.

G. Baker, Department of Statistics
University of South Carolina; Slide 46



Standard Deviations	–3	–2	–1	0	+1	+2	+3
Cumulati Percentages	.1	2.3	15.9	50	84.1	97.7	99.9
Z Scores	%	%	%	%	%	%	%
T Scores	–3.	–2.	–1.0	0	+1.0	+2.0	+3.0
SAT Scores	0	0	40	50	60	70	80
Spelling Scores	20	30	400	500	600	700	800
IQ Scores	20	300	8	11	14	17	20



99.7%
95%
68%
Mean
Number of people with each score
200 300 400 500 600 700 800
Score on Scholastic Assessment Test
−3 −2 −1 0 +1 +2 +3
Standard deviation
99.9 98 84 50 16 2 0.1
Percent who exceed score

A Note about Transformations:

Adding or subtracting just shifts the distribution, without changing the variation (variance).

Multiplying or dividing changes the variability, but it is a multiple of the transformation

Transformations can be useful when scores are skewed from the “mean” that the researcher prefers to work with. Also, transformations can be helpful when comparing multiple sets of scores.

Standard Deviation: Example

X	$\left(X - \overline{X}\right)$	$\left(X - \overline{X}\right)^2$
21	-5.8	33.64
25	-1.8	3.24
24	-2.8	7.84
30	3.2	10.24
34	7.2	51.84
26.8	**0**	**106.8**

$$S = \sqrt{\frac{106.8}{5}} = \sqrt{21.36} = 4.62$$

A note on *N vs. n* and *N vs. n+1*

N is used to represent the number of scores in a population. Some authors also use it to represent the number of scores in a sample as well.

n is used to represent the number of scores in a sample.

When calculating the variance and standard deviation, you will sometimes see **N + 1** or **n + 1** instead of **N** or **n**.

N+1 or **n+1** is used by many social scientists when computing from a sample **with the intention of generalizing to a larger population**. **N+1** provides a better estimate and is the formula used by SPSS.

$\hat{S}$ Estimating the population standard deviation from a sample

S, the sample standard, is usually a little smaller than the population standard deviation. Why?

The sample mean minimizes the sum of squared deviations (SS). Therefore, if the sample mean differs at all from the population mean, then the SS from the sample will be an ***underestimate*** of the SS from the population

Therefore, statisticians alter the formula of the sample standard deviation by subtracting 1 from *N*

Variance is Error in Predictions

The larger the variability, the larger the differences between the mean and the scores, so the larger the error when we use the mean to predict the scores

Error or error variance: average error between the predicted mean score and the actual raw scores

Same for the population: estimate of population variance

Summarizing Research Using Variability

Remember, the standard deviation is most often the measure of variability reported.

The more consistent the scores are (i.e., the smaller the variance), the stronger the relationship.

Chapter 4: Review

Review: Measures of Variability

Just what is variability?

Variability is the **spread or dispersion of scores**.

Measuring Variability

There are a few ways to measure variability and they include:

1) **The Range**
2) **The Mean Deviation**
3) **The Standard Deviation**
4) **The Variance**

Review: The Range

Measures of Variability

Range: The range is a measure of the distance between highest and lowest.

R= H – L

Temperature Example:

		Range:
Honolulu: 89° – 65°		24°
Phoenix: 106° – 41°	65°	

Review: Variance and Standard Deviation

Variance: is a **measure of the dispersion of a sample** (or how closely the observations cluster around the mean [average])

Standard Deviation: the **square root of the variance**, is the measure of variation in the observed values (or variation in the clustering around the mean).

Review: The Variance

Remember that the **deviation** is the distance of any given score from its mean.

$$(X - \overline{X})$$

The **variance** takes into account **every** score.

But if we were to simply add them up, the plus and minus (positive and negative) scores would cancel each other out because the sum of actual deviations is always zero!

$$\sum (X - \overline{X}) = 0$$

Review: The Variance

So, what we should we do?

We **square** the actual deviations and then **add** them together.

$$\sum (X - \bar{X})^2$$

– Remember: When you square a negative number it becomes positive!

SO,

S^2 = sum of squared deviations divided by the number of scores.

The variance provides information about the relative variability.

Review: Variance: Weeks on Unemployment:

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev		
X (weeks)	Deviation: $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)		
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$		
$\Sigma X=30$ $\bar{X} = \frac{30}{6} = 5$		$\sum (X - \bar{X})^2 = 52$		

Review: The Variance

The mean of the squared deviations is the same as the variance, and can be symbolized by s^2

$$s^2 = \frac{\sum\left(X - \overline{X}\right)^2}{N}$$

where $s^2 =$ variance

$\sum(X - \overline{X})^2 =$ sum of the squared deviations from the mean

$N =$ total number of scores

Review: Variance: Weeks on Unemployment:

Step 1: Calculate the Mean	Step 2: Calculate Deviation	Step 3: Calculate Sum of square Dev	Step 4: Calculate the Mean of squared dev.

X (weeks)	**Deviation:** $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)	**Variance:** $s^2 = \frac{\sum (X - \bar{X})^2}{N}$	
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$	$\frac{52}{6} = 8.67$ (weeks squared)	
$\Sigma X=30$ $\bar{X} = \frac{30}{6} = 5$		$\sum (X - \bar{X})^2 = 52$		

Review: Standard Deviation

Standard Deviation:

It is the typical **(standard)** difference **(deviation)** of an observation from the mean.

Think of it as the average distance a data point is from the mean, although this is not strictly true.

Review: Standard Deviation?

Standard Deviation:

The standard deviation is calculated by taking **the square root of the variance**.

$$s = \sqrt{\sum \frac{(X - \bar{X})^2}{n}}$$

Review: Standard Deviation: Weeks on Unemployment:

Step 1: Calculate the Mean

Step 2: Calculate Deviation

Step 3: Calculate Sum of square Dev

Step 4: Calculate the Mean of squared dev.

Step 5: Calculate the Square root of the Var.

X (weeks)	**Deviation:** $(X - \bar{X})$	$(X - \bar{X})^2$ (raw score from the mean, squared)	**Variance:** $s^2 = \frac{\sum(X - \bar{X})^2}{N}$	**Standard Deviation:** $s = \sqrt{\sum\frac{(X - \bar{X})^2}{n}}$ (square root of the variance)
9 8 6 4 2 1	9-5= 4 8-5=3 6-5=1 4-5=-1 2-5=-3 1-5=-4	$4^2 = 16$ $3^2 = 9$ $1^2 = 1$ $-1^2 = 1$ $-3^2 = 9$ $-4^2 = 16$	$\frac{52}{6} = 8.67$ (weeks squared)	$\sqrt{8.67}$
$\Sigma X=30$ $\bar{X} = \frac{30}{6} = 5$		$\sum(X - \bar{X})^2 = 52$		s = 2.94

THANK YOU