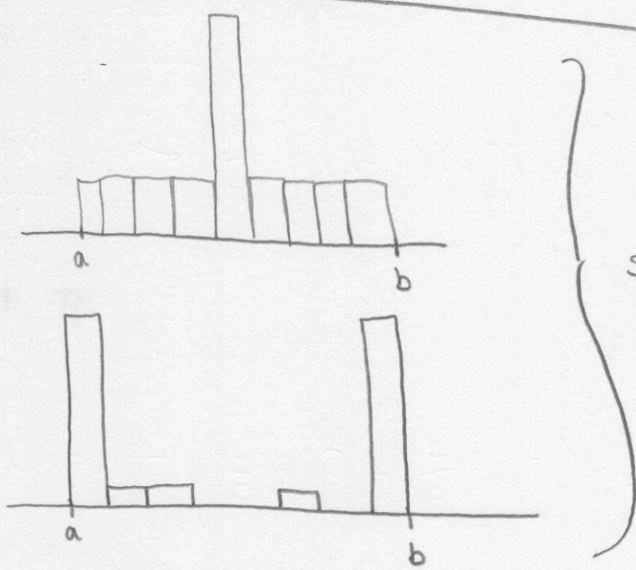
§ 2.3 MEASURES OF VARIABILITY (VARIATION)

1. RANGE : MAX − MIN
2. VARIANCE
3. STANDARD DEVIATION

RANGE = MAX − MIN

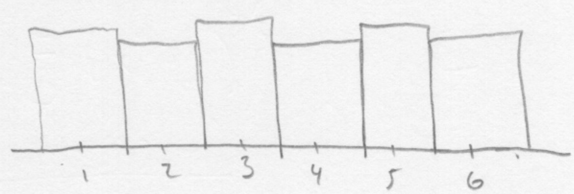
CONSIDER 2 DISTRIBUTIONS:



SAME RANGE

ROLL A FAIR DIE 100 TIME:

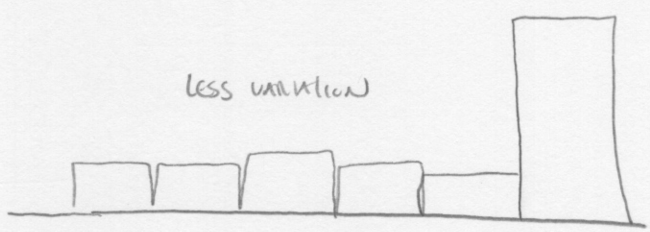
MORE VARIATION



ROLL A WEIGHTED DIE

$\underbrace{\text{WEIGHTED DIE}}$ COMES UP "6" MORE OFTEN

LESS VARIATION



MORE VARIATION = LESS PREDICTABLE

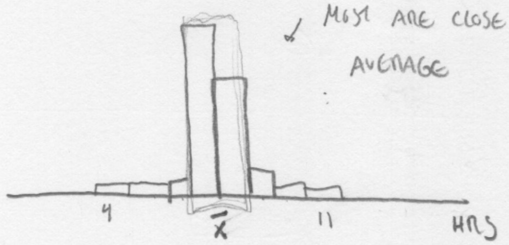
LESS VARIATION = MORE PREDICTABLE

COMPANY A, COMPANY B BOTH MAKE BATTERIES.

LIFETIME OF BATTERIES FOR EACH HAS DISTRIBUTION;

LESS VARIATION

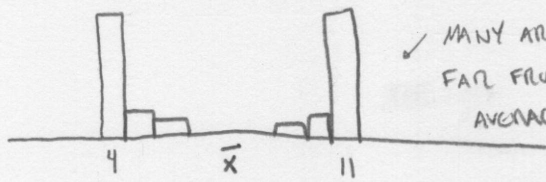
A:



MOST ARE CLOSE TO AVERAGE

MORE VARIATION

B



MANY ARE FAR FROM AVERAGE

BOTH COMPANIES HONESTLY CLAIM THEIR BATTERIES LAST AN AVERAGE OF 7.5 HRS.

"QUALITY CONTROL"

(LOW VARIATION)

HOW DO WE QUANTATATIVELY MEASURE VARIATION?

2. VARIANCE: FOR A POPULATION $\quad \Sigma \quad \sigma$

$$\text{VARIANCE } \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

FOR A SAMPLE

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

POPULATION, $n = 6$

ex. DATA SET: $\underset{x_1}{3}, \underset{x_2}{4}, \underset{x_3}{6}, \underset{x_4}{10}, \underset{x_5}{11}, \underset{x_6}{26}$

$$\text{MEAN } \mu = \frac{\sum x_i}{n} = \frac{3+4+\dots+26}{6} = \frac{60}{6} = 10$$

COMPUTING THE VARIANCE:

x_i	$x_i - \mu$	$(x_i - \mu)^2 \geq 0$
3	-7	49
4	-6	36
6	-4	16
10	0	0
11	1	1
26	16	+ 256

(SUM 0)

$$\sum (x_i - \mu)^2 = 358$$

VARIANCE IS AVERAGE OF THESE #'S.

BY DEFAULT: 4 DECIMAL PLACES.

$$\text{VARIANCE } \sigma^2 = \frac{\sum (x_i - \mu)^2}{n} = \frac{358}{6} = \underline{\underline{59.6667}}$$

UNITS FOR VARIANCE:

ex. MEASUREMENTS HAVE UNIT OF <u>HOURS</u>.

VARIANCE HAS UNIT HOURS^2 (?!)

2.5
↓

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2 \geq 0$
1.3	-1.2	1.44
2.5	0	0
2.8	.3	.09
3.4	.9	+ .81
	TOTAL 0 ✓	2.34

VARIANCE (SAMPLE)

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

$$s^2 = \frac{2.34}{4-1} = .78$$

VARIANCE

STAND DEV $\quad s = \sqrt{s^2} = \sqrt{.78} = \boxed{.8832}$

SAMPLE STATISTICS	POPULATION PARAMETERS
SAMPLE SIZE n	POPULATION SIZE N
SAMPLE MEAN $\bar{x}$	POPULATION MEAN μ
SAMPLE VARIANCE s^2	POPULATION VARIANCE σ^2
$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ DIV. BY SMALLER #, RESULT IS BIGGER; MORE ACCURATE ESTIMATE FOR σ^2	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
SAMPLE STAND. DEV. s	POPULATION STAND. DEV. σ
$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$

(POPULATION HAS MORE VARIATION THAN THE SAMPLE)

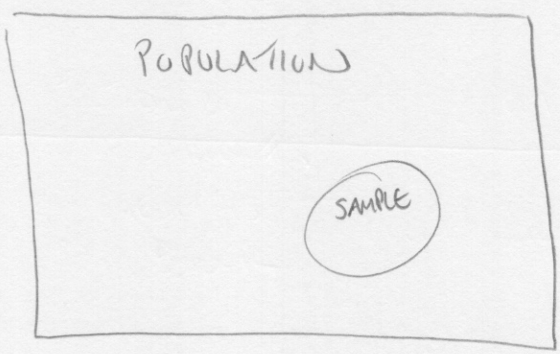
Def: THE POPULATION IS THE ENTIRE SET OF THINGS YOU ARE INTERESTED IN.

THE SAMPLE IS THE SUBSET OF THE POPULATION THAT YOU ACTUALLY STUDY.

ex. PREDICTIONS FOR AN ELECTION.

POPULATION : EVERYONE WHO'S GOING TO VOTE.

SAMPLE : SOME OF THE PEOPLE WHO ARE GOING TO VOTE



EX 2.17 1.28 2.39 1.50 1.88 1.51

RANGE: MAX - MIN = 2.39 - 1.28 = $\boxed{1.11}$

VARIANCE: σ^2 s^2 $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

MEAN $\bar{X} = \dfrac{\sum x_i}{n} = \dfrac{1.28 + \ldots + 1.51}{5} = \dfrac{8.56}{5}$

$\bar{x} = 1.712$ $n = 5$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1.28	-.432	.1866
2.39	.678	.4600
1.50	-.212	.0449
1.88	.168	.0282
1.51	-.202	.0408

-1.712 SQUARED

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

$= \dfrac{.1866 + \ldots + .0408}{5-1}$

$= \dfrac{.7605}{4} = .1901$

VAR. $s^2 = .1901$

STAND. DEV. $s = \sqrt{s^2} = \sqrt{.1901} = \boxed{.4360}$

§2.4 ON THE PRACTICAL SIGNIFICANCE OF THE STANDARD DEVIATION

CHEVYCHEV'S THM:

GIVEN A NUMBER $k \geq 1$, AND A SET OF n MEASUREMENTS, AT LEAST

$$\underbrace{\left[\, 1 - \frac{1}{k^2} \,\right]}_{\text{PROPORTION}}$$

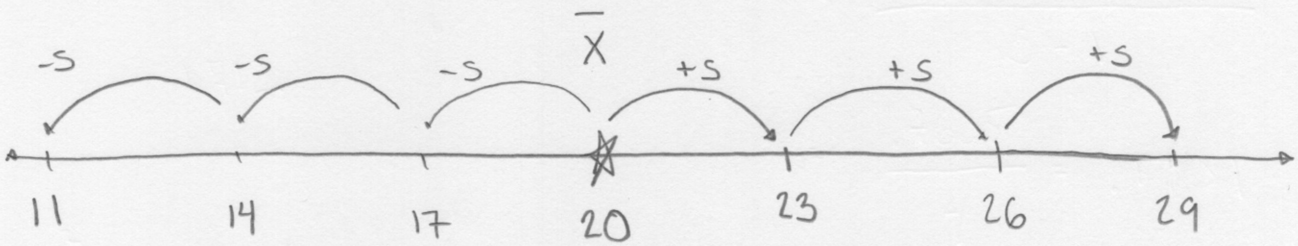
OF THE MEASUREMENTS LIE WITHIN k STANDARD DEVIATIONS OF THE MEAN.

☆ HOME BASE ON # LINE : MEAN

STEP-SIZE : STAND. DEV.

ex. MEAN 20

STAND. DEV. 3



$\bar{x} - 2s$ $\bar{x} + 2s$

#'S THAT LIE WITHIN 1 S.D. OF MEAN: 17 to 23

14 LIE WITHIN 2 S.D. OF MEAN 26

ex. $\bar{x} = 110$ DESCRIBE THE SET OF VALUES THAT LIE

$s = 25$ WITH

(a) 1 S.D. OF THE MEAN

(b) 2 S.D. OF THE MEAN

(c) 3 S.D. OF THE MEAN

(a) $\bar{x} - s$ up to $\bar{x} + s$

$110 - 25$ $\qquad$ $110 + 25$

85 $\qquad$ 135

$85 \leq x \leq 135$

(b) $\bar{x} - 2s$ up to $\bar{x} + 2s$

$110 - 2(25)$ $\qquad$ $110 + 2(25)$

$60 \leq x \leq 160$

(c) $35 \leq x \leq 185$

$k = 1$: AT LEAST $\left[1 - \frac{1}{1^2}\right] = 0$ MEASUREMENTS

LIE WITHIN 1 S.D. OF THE MEAN.

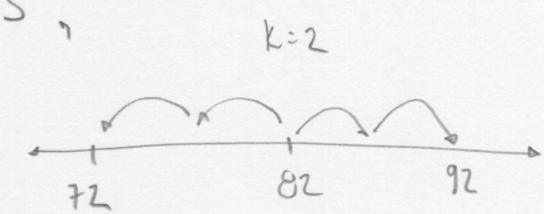
$k = 2$: AT LEAST $\left[1 - \frac{1}{2^2}\right] = 1 - \frac{1}{4} = \frac{3}{4}$

OF THE MEASUREMENTS LIE WITHIN 2 S.D. OF MEAN.

1000 MEASUREMENTS

$\bar{x} = 82$

$s = 5$



AT LEAST $\frac{3}{4}$ (75%) OF MEASUREMENTS &

LIE BETWEEN 72 $\boxed{\&}$ 92. & &

AND

$k = 3$ AT LEAST $\left[1 - \frac{1}{3^2}\right] = 1 - \frac{1}{9} = \frac{8}{9} = .8888$

88.88% OF MEASUREMENTS LIE WITHIN

3 S.D. OF MEAN.

DECIMAL → PERCENT

MOVE DECIMAL POINT

2 SPACES TO THE RIGHT

ex. $.30. = 30\%$

$.0003 = .03\%$

$1.321 \rightarrow 132.1\%$