DESCRIPTIVE MEASURES

Measures of Variability or Dispersion

Why VARIABILITY matters?

Data Set 1:

80, 85, 85, 90, 95

mean = 87

Data Set 2:

25, 65, 70, 125, 150

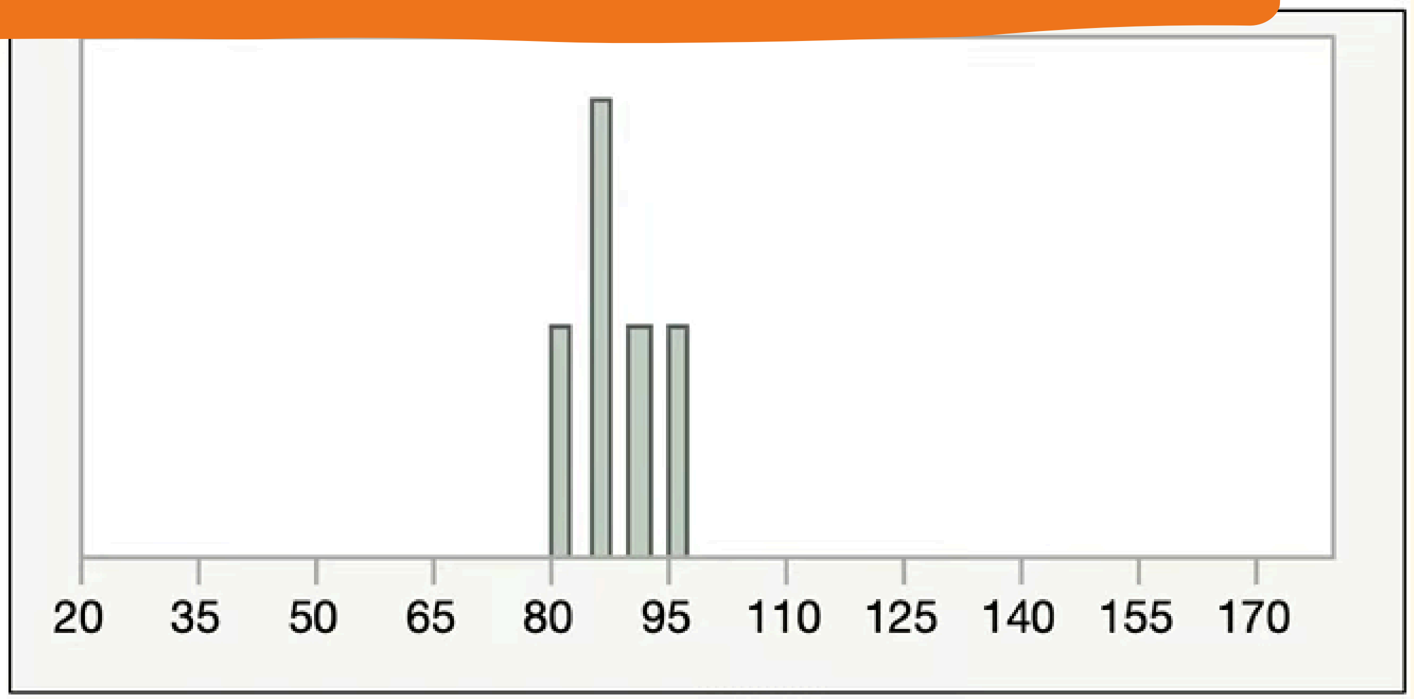
mean = 87

Why VARIABILITY matters?

Data Set 1:

80, 85, 85, 90, 95

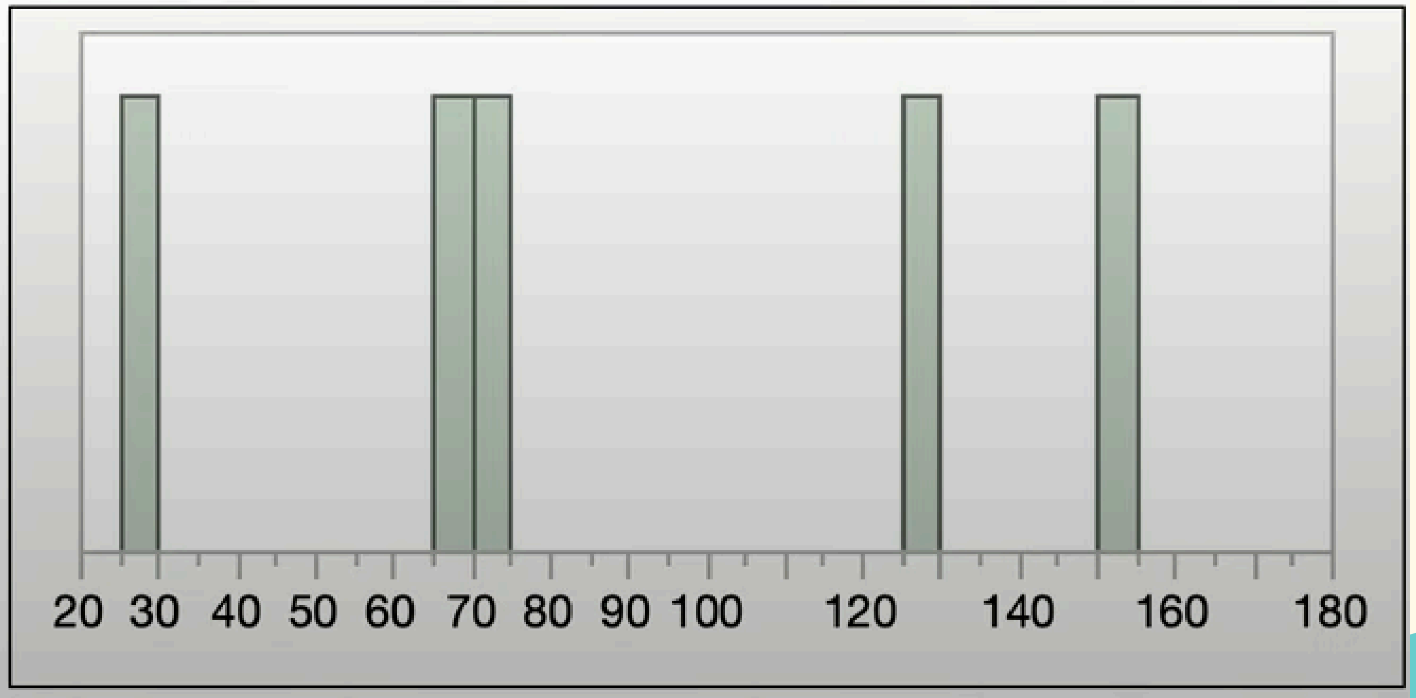
mean = 87



Data Set 2:

25, 65, 70, 125, 150

mean = 87

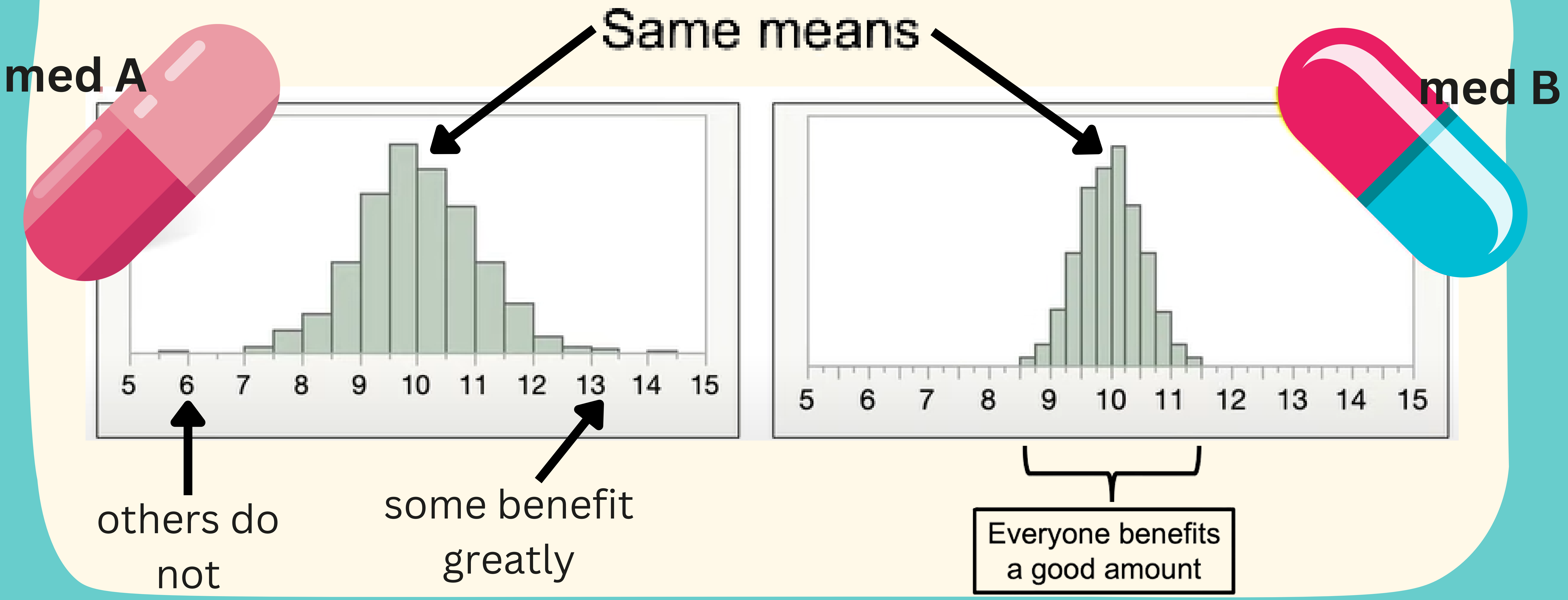


Why VARIABILITY matters?

med A

med B

Why VARIABILITY matters?



Why VARIABILITY matters?

A testing lab wishes to test two experimental brands of outdoor paint to see how long each will last before fading. The testing lab makes 6 gallons of each paint to test. Since different chemical agents are added to each group and only six cans are involved, these two groups constitute two small populations. The results (in months) are shown. Find the mean of each group.

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

Why VARIABILITY matters?

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

Mean for Brand A:

$$\mu = \frac{\sum x}{N} = \frac{210}{6} = 35\ months$$

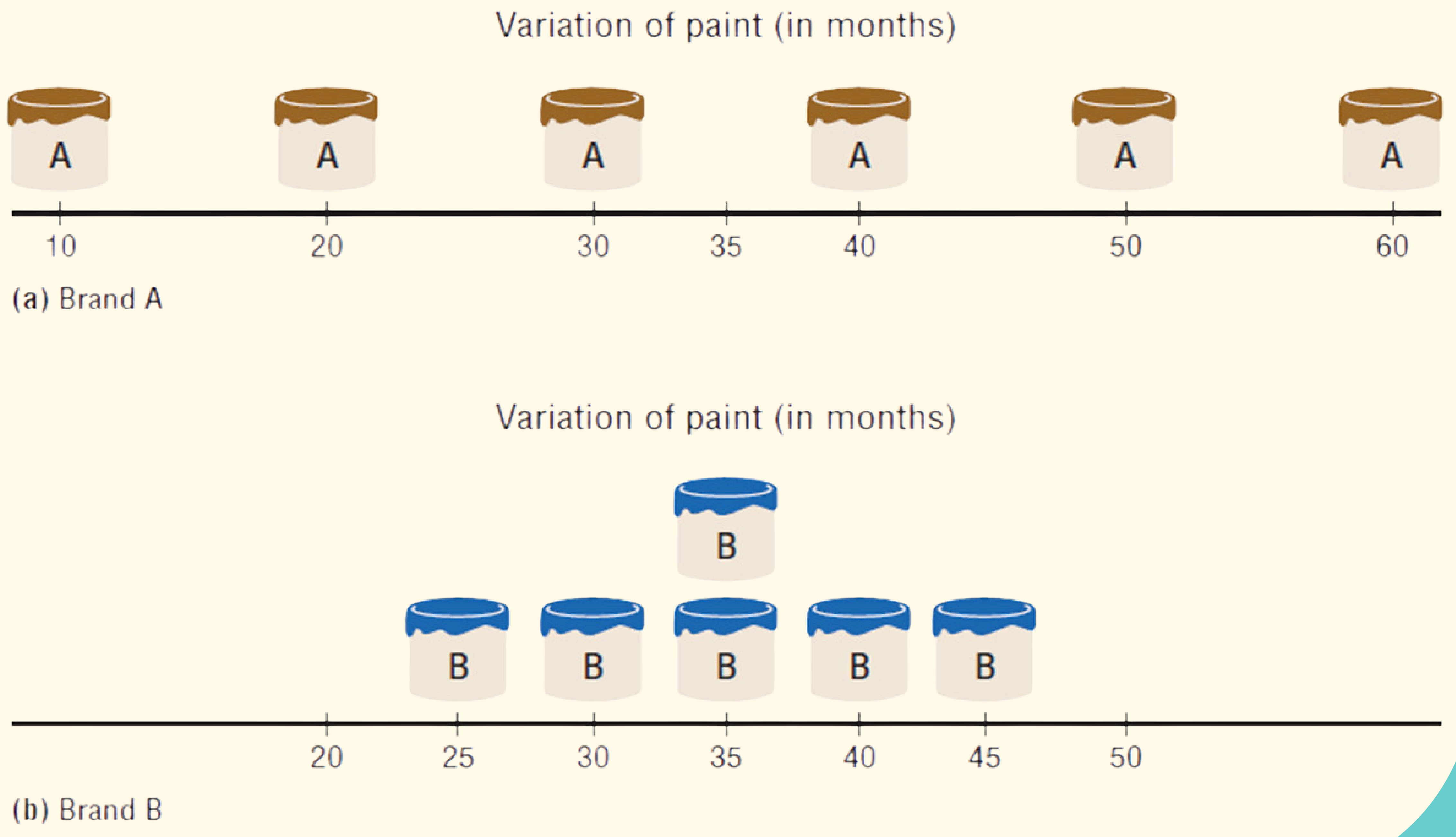
Mean for Brand B:

$$\mu = \frac{\sum x}{N} = \frac{210}{6} = 35\ months$$

Since the means are equal, can we conclude that both brands of paint last equally well?

Why VARIABILITY matters?

On the right is a graphical representation of the data set. What can you say about the two brands of paint based on the graphs?



Measures of VARIABILITY

How can we describe these differences statistically?

In statistics, **measures of variability** describe how data values in a given data set differ from one another (e.g. the spread or clustering of points)

Range

Variance

Standard Deviation

RANGE

The **range** is the simplest measure of variability (or dispersion), and is defined as follows:

$$R = HV - LV$$

The range

The **h**ighest score in the set

The **l**owest score in the set

Calculating the RANGE

Data Set 1:

80, 85, 85, 90, 95

mean = 87

R = 95 - 80 = 15

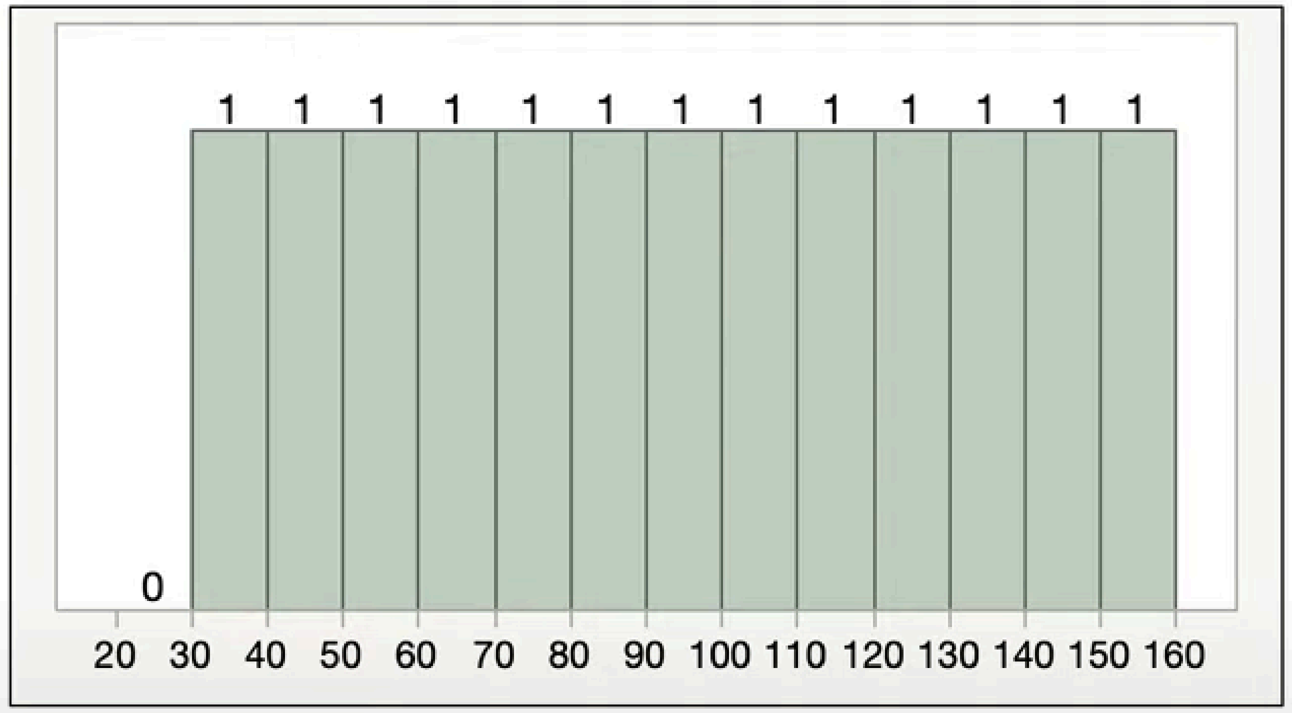
Data Set 2:

25, 65, 70, 125, 150

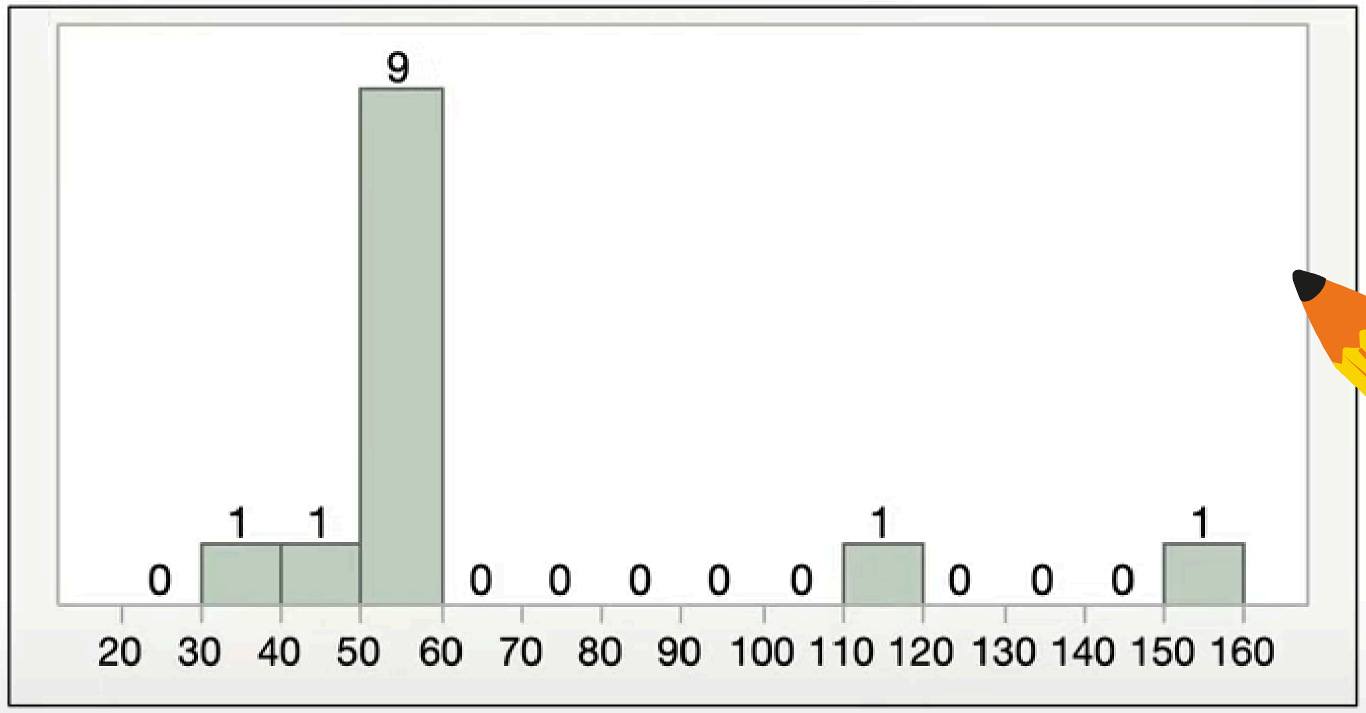
mean = 87

R = 150 - 25 = 125

SAME range, DIFFERENT distribution



$R = 150 - 30 = 120$



$R = 150 - 30 = 120$

STANDARD DEVIATION

The standard (or typical) amount that data **deviate** from the mean.

Populations	Samples
σ	s
“sigma”	“s”

VARIANCE

The averaged **squared** deviation from the mean (the square of standard deviation)

Populations	Samples
σ^2	s^2
“sigma squared”	“s squared”

FORMULA

	Population Parameter:	Sample Statistic:
Standard Deviation:	$\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$
Variance:	$\sigma^2 = \dfrac{\sum(x-\mu)^2}{N}$	$s^2 = \dfrac{\sum(x-\bar{x})^2}{(n-1)}$

EXAMPLE 1

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

A testing lab wishes to test two experimental brands of outdoor paint to see how long each will last before fading. The testing lab makes 6 gallons of each paint to test. Since different chemical agents are added to each group and only six cans are involved, these two groups constitute two small populations. The results (in months) are shown on the left.

Example 1 SOLUTION:

Brand A: 10, 60, 50, 30, 40, 20

$$\mu = \frac{\sum x}{N} = \frac{210}{6} = 35\ months$$

R = 60 - 10 = 50 months

x	$(x-\mu)$	$(x-\mu)^2$
10	**10 – 35 = –25**	**625**
60	**60 – 35 = 25**	**625**
50	**50 – 35 = 15**	**225**
30	**30 – 35 = –5**	**25**
40	**40 – 35 = 5**	**25**
20	**20 – 35 = –15**	**225**
$\sum(x-\mu)^2$		**1,750**

Brand B: 35, 45, 30, 35, 40, 25

$$\mu = \frac{\sum x}{N} = \frac{210}{6} = 35\ months$$

R = 45 - 25 = 20 months

x	$(x-\mu)$	$(x-\mu)^2$
35	**35 – 35 = 0**	**0**
45	**45 – 35 = 10**	**100**
30	**30 – 35 = –5**	**25**
35	**35 – 35 = 0**	**0**
40	**40 – 35 = 5**	**25**
25	**25 – 35 = –10**	**100**
$\sum(x-\mu)^2$		**250**

Example 1 SOLUTION:

Brand A: 10, 60, 50, 30, 40, 20

x	$(x-\mu)$	$(x-\mu)^2$
10	**10 – 35 = –25**	**625**
60	**60 – 35 = 25**	**625**
50	**50 – 35 = 15**	**225**
30	**30 – 35 = –5**	**25**
40	**40 – 35 = 5**	**25**
20	**20 – 35 = –15**	**225**
$\sum(x-\mu)^2$		**1,750**

Population VARIANCE:

$$\sigma^2 = \frac{1750}{6} = \frac{875}{3}$$

Population STANDARD DEVIATION:

$$\sigma = \sqrt{875/3} = 17.08$$

Brand B: 35, 45, 30, 35, 40, 25

x	$(x-\mu)$	$(x-\mu)^2$
35	**35 – 35 = 0**	**0**
45	**45 – 35 = 10**	**100**
30	**30 – 35 = –5**	**25**
35	**35 – 35 = 0**	**0**
40	**40 – 35 = 5**	**25**
25	**25 – 35 = –10**	**100**
$\sum(x-\mu)^2$		**250**

Population VARIANCE:

$$\sigma^2 = \frac{250}{6} = \frac{125}{3}$$

Population STANDARD DEVIATION:

$$\sigma = \sqrt{125/3} = 6.45$$

Example 1 SOLUTION:

Brand A: 10, 60, 50, 30, 40, 20

Population VARIANCE:

$$\sigma^2 = \frac{1750}{6} = \frac{875}{3}$$

Population STANDARD DEVIATION:

$$\sigma = \sqrt{875/3} = 17.08$$

Brand B: 35, 45, 30, 35, 40, 25

Population VARIANCE:

$$\sigma^2 = \frac{250}{6} = \frac{125}{3}$$

Population STANDARD DEVIATION:

$$\sigma = \sqrt{125/3} = 6.45$$

Brand A and Brand B have the same mean,
but different range and standard deviation.

This concludes that Brand B's data are closely related to each other.
Thus, Brand A is more variable.

EXAMPLE 2

Let's have a look and compare the scores Sophia and Daff in the recent 4 summative tests by solving the mean, range, variance and standard deviation.

Sofia's Scores: 75, 75, 75, 100

Daff's Scores: 80, 81, 82, 82

Example 2 SOLUTION:

Sofia's Scores: 75, 75, 75, 100

$$\bar{x} = \frac{75 + 75 + 75 + 100}{4} = \frac{325}{4} = 81.25$$

R = 100 - 75 = 25

x	$(x - \bar{x})$	$(x - \bar{x})^2$
75	75 – 81.25 = –6.25	39.0625
75	75 – 81.25 = –6.25	39.0625
75	75 – 81.25 = –6.25	39.0625
100	100 – 81.25 = 18.75	351.5625
$\sum(x - \bar{x})^2$		468.75

Daff's Scores: 80, 81, 82, 82

$$\bar{x} = \frac{80 + 81 + 82 + 82}{4} = \frac{325}{4} = 81.25$$

R = 82 - 80 = 2

x	$(x - \bar{x})$	$(x - \bar{x})^2$
80	80 – 81.25 = –1.25	1.5625
81	81 – 81.25 = –0.25	0.0625
82	82 – 81.25 = 0.75	0.5625
82	82 – 81.25 = 0.75	0.5625
$\sum(x - \bar{x})^2$		2.75

Example 2 SOLUTION:

Sofia's Scores: 75, 75, 75, 100

x	$(x-\overline{x})$	$(x-\overline{x})^2$
75	75 – 81.25 = –6.25	39.0625
75	75 – 81.25 = –6.25	39.0625
75	75 – 81.25 = –6.25	39.0625
100	100 – 81.25 = 18.75	351.5625
$\sum(x-\overline{x})^2$		468.75

Sample VARIANCE:

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{468.75}{3} = 156.25$$

Sample STANDARD DEVIATION:

$$s = \sqrt{156.25} = 12.50$$

Daff's Scores: 80, 81, 82, 82

x	$(x-\overline{x})$	$(x-\overline{x})^2$
80	80 – 81.25 = –1.25	1.5625
81	81 – 81.25 = –0.25	0.0625
82	82 – 81.25 = 0.75	0.5625
82	82 – 81.25 = 0.75	0.5625
$\sum(x-\overline{x})^2$		2.75

Sample VARIANCE:

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{2.75}{3}$$

Sample STANDARD DEVIATION:

$$s = \sqrt{(2.75)/3} = 0.96$$

Example 2 SOLUTION:

Sofia's Scores: 75, 75, 75, 100

Sample VARIANCE:

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{468.75}{3} = 156.25$$

Sample STANDARD DEVIATION:

$$s = \sqrt{156.25} = 12.50$$

Daff's Scores: 80, 81, 82, 82

Sample VARIANCE:

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{2.75}{3}$$

Sample STANDARD DEVIATION:

$$s = \sqrt{(2.75)/3} = 0.96$$

Sofia and Daff have the same mean,
but different range and standard deviation.

This conclude that Daff's scores are closely related to each other.
Thus, in general Daff is consistent with her performance.