Chapter 4: Variability

Variability

- Provides a quantitative measure of the degree to which scores in a distribution are spread out or clustered together

Central Tendency and Variability

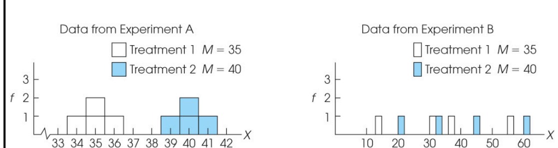
- Central tendency describes the central point of the distribution, and variability describes how the scores are scattered around that central point.
- Together, central tendency and variability are the two primary values that are used to describe a distribution of scores.

Variability

- Variability serves both as a descriptive measure and as an important component of most inferential statistics.
- As **a descriptive statistic**, variability measures the degree to which the scores are spread out or clustered together in a distribution.
- In the context of **inferential statistics**, variability provides a measure of how accurately any individual score or sample represents the entire population.

Variability (cont.)

- When the population variability is small, all of the scores are clustered close together and any individual score or sample will necessarily provide a good representation of the entire set.
- On the other hand, when variability is large and scores are widely spread, it is easy for one or two extreme scores to give a distorted picture of the general population.



Measuring Variability

- Variability can be measured with
  - the range
  - the interquartile range
  - the standard deviation/variance.
- In each case, variability is determined by measuring *distance*.

The Range

- The **range** is the total distance covered by the distribution, from the highest score to the lowest score (using the upper and lower real limits of the range).

Range

- URL x_{max} - LRL x_{min}
  - e.g. 3, 7, 12, 8, 5, 10

Problems?

- Distribution 1
  - 1, 8, 9, 9, 10, 10 R = ?
- Distribution 2
  - 1, 2, 3, 6, 8, 10 R = ?

The Interquartile Range

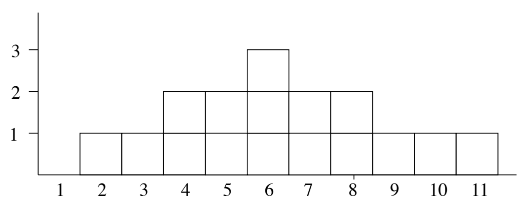
- The **interquartile range** is the distance covered by the middle 50% of the distribution (the difference between Q1 and Q3).

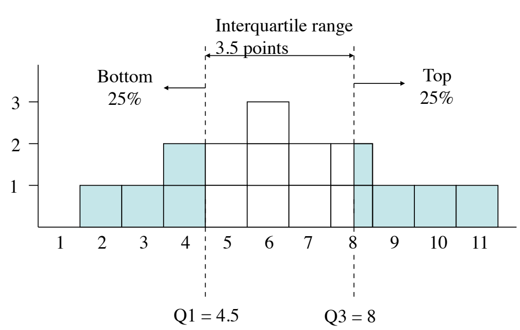
Scores

2, 3, 4, 4, 5, 5, 6, 6,

6, 7, 7, 8, 8, 9, 10, 11

x	f	cf	cp	c%
11	1	16	16/16	100%
10	1	15	15/16	93.75%
9	1	14	14/16	87.5%
8	2	13	13/16	81.25%
7	2	11	11/16	68.75%
6	3	9	9/16	56.25%
5	2	6	6/16	37.5%
4	2	4	4/16	25%
3	1	2	2/16	12.5%
2	1	1	1/16	6.25%



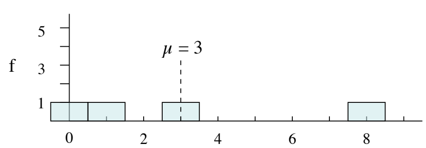


The Standard Deviation

- **Standard deviation** measures the **standard** (or average) distance between a score and the mean.

0, 1, 3, 8 $\quad \mu = \frac{8+1+3+0}{4} = 3$

x	(x- μ)
8	8 - 3 = +5
1	1 - 3 = -2
3	3 - 3 = 0
0	0 - 3 = -3



---

x	x - μ	(x - μ)^2
1	1 - 2 = -1	1
0	0 - 2 = -2	4
6	6 - 2 = +4	16
1	1 - 2 = -1	1

$\sum x = 8$

$\mu = 2$

$22 = \sum(x - \mu)^2 = SS$

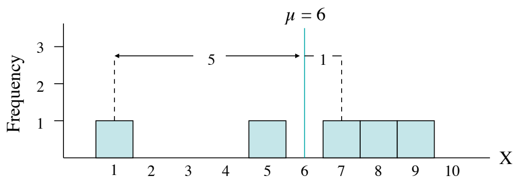
or

x	x^2
1	1
0	0
6	36
1	1

$\sum x = 8$

$\sum x^2 = 38$

$$SS = \sum x^2 - \frac{(\sum x)^2}{N}$$
$$= 38 - \frac{8^2}{4}$$
$$= 38 - 16$$
$$= 22$$



- 1, 9, 5, 8, 7
- $\mu = 6$

x	(x - μ)	(x - μ)^2
1	1 - 6 = -5	25
9	9 - 6 = +3	9
5	5 - 6 = -1	1
8	8 - 6 = +2	4
7	7 - 6 = +1	1

$$\sum(x - \mu)^2 = 40 = SS$$

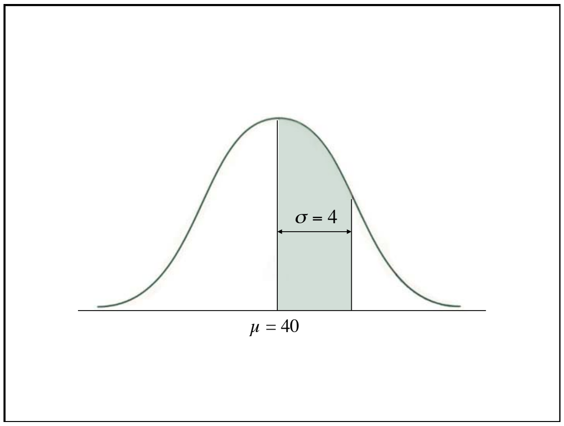
$$\sigma^2 = \frac{SS}{N} = \frac{\sum(x - \mu)^2}{N} = \frac{40}{5} = 8$$

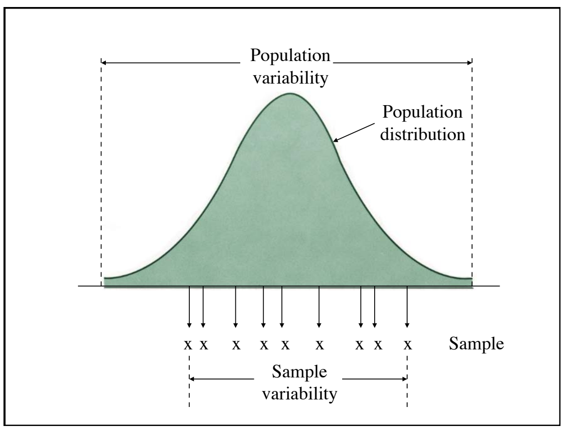
$$\sigma = \sqrt{\frac{SS}{N}} = \sqrt{\frac{\sum(x - \mu)^2}{N}} = 2.83$$

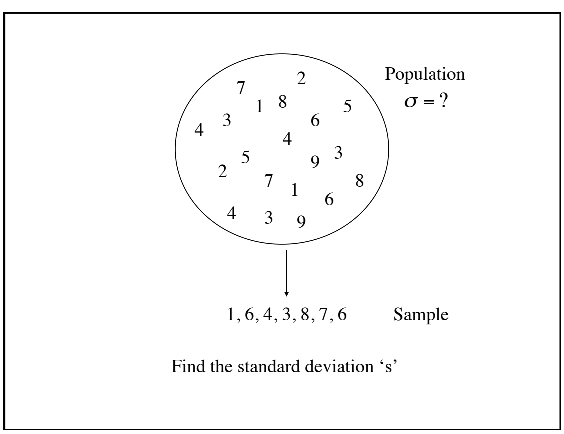
Variance and Standard Deviation for a population of scores

$$\sigma^2 = \frac{SS}{N} = \frac{\sum(x - \mu)^2}{N}$$

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


$\sigma = 4$
$\mu = 40$


Population variability
Population distribution
x x x x x x x x x
Sample
Sample variability


7 2 1 8 5 4 3 6 4 5 9 3 2 7 1 8 6 4 3 9
Population
$\sigma = ?$
1, 6, 4, 3, 8, 7, 6
Sample
Find the standard deviation 's'

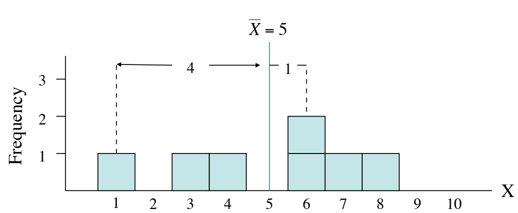
Variance and Standard Deviation for a Sample Used to Estimate the Population Value

Variance:

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

$$s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{n-1}}$$

1, 6, 4, 3, 8, 7, 6,



1, 6, 4, 3, 8, 7, 6

Sample

$\overline{X} = \frac{\sum x}{n} = \frac{35}{7} = 5$

x	$(x-\overline{X})$	$(x-\overline{X})^2$
1	1 - 5 = -4	16
6	6 - 5 = +1	1
4	4 - 5 = -1	1
3	3 - 5 = - 2	4
8	8 - 5 = +3	9
7	7 - 5 = +2	4
6	6 - 5 = +1	1

$\sum(x-\overline{X})^2 = SS = 36$

variance $s^2 = \frac{\sum(x-\overline{X})^2}{n-1}$ or $\frac{SS}{n-1}$

standard deviation $s = \sqrt{\frac{\sum(x-\overline{X})^2}{n-1}} = \sqrt{\frac{36}{6}} = \sqrt{6} = 2.45$ or $\sqrt{\frac{SS}{n-1}}$

Sum of Squares

$$SS = \sum(x-\overline{X})^2$$

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

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

But Also:

$$SS = \sum x^2 - \frac{(\sum x)^2}{n}$$

x	x^2
1	1
6	36
4	16
3	9
8	64
7	49
6	36
35	211

$$SS = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$= 211 - \frac{35^2}{7}$$

$$= 211 - \frac{1225}{7}$$

$$= 211 - 175 \quad = 36$$

$$\sigma^2 = \frac{SS}{N} = \frac{\sum(x-\mu)^2}{N}$$

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

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

$$s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{\sum(x-\overline{X})^2}{n-1}}$$

Example

- Randomly select a score from a population

  x = 47

- What value would you predict for the population mean?

  if $\sigma = 4$

  if $\sigma = 20$

Properties of the Standard Deviation

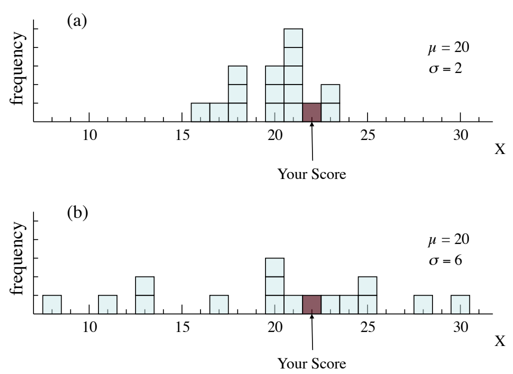
1. The same score can have very different meanings in 2 different distributions
2. Standard deviation helps us make predictions about sample data

   e.g. Figure 4.8 — low variability / high variability

   What is the probability of picking a score near μ = 20 ?

3. Sampling error - how big?

   (standard deviation a measure)



Transformations of Scale

1. Adding a constant to each score will not change the standard deviation
2. Multiplying each score by a constant causes the standard deviation to be multiplied by the same constant

Comparing Measures of Variability

- Two considerations determine the value of any statistical measurement:

1. The measures should provide a stable and reliable description of the scores. It should not be greatly affected by minor details in the set of data.
2. The measure should have a consistent and predictable relationship with other statistical measurements.

Factors that Affect Variability

1. Extreme scores
2. Sample size
3. Stability under sampling
4. Open-ended distributions

Relationship with Other Statistical Measures

- Variance and standard deviation are mathematically related to the mean. They are computed from the squared deviation scores (squared distance of each score from the mean).
- Median and semi-interquartile range are both based on percentiles and therefore are used together. When the median is used to report central tendency, semi-interquartile range is often used to report variability.
- Range has no direct relationship to any other statistical measure.

Sample variability and degrees of freedom

df = n - 1

The Mean and Standard Deviation as Descriptive Statistics

- If you are given numerical values for the mean and the standard deviation, you should be able to construct a visual image (or a sketch) of the distribution of scores.
- As a general rule, about 70% of the scores will be within one standard deviation of the mean, and about 95% of the scores will be within a distance of two standard deviations of the mean.

Mean number of errors on easy vs. difficult tasks for males vs. females

	Easy	Difficult
Female	1.45	8.36
Male	3.83	14.77

41

When we report descriptive statistics for a sample, we should report a measure of central tendency and a measure of variability.

Mean number of errors on easy vs. difficult tasks for males vs. females

	Easy	Difficult
Female	M =1.45 SD = .92	M = 8.36 SD = 2.16
Male	M =3.83 SD =1.24	M =14.77 SD = 3.45

43